# Centrality Analysis for Modified Lattices

Martin W. Paton
Department of Mathematics & Statistics
University of Strathclyde
Glasgow, U.K.

2022

This thesis is submitted to the University of Strathclyde for the degree of Doctor of Philosophy in the Faculty of Science.

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material in, or derived from, this thesis.

Signed:

Date:

# Acknowledgements

I would like to thank the University of Strathclyde Strategic Technology Partnership with CAPITA for providing the funding opportunity to pursue this research. Without this financial support I would not have had this amazing opportunity.

I will forever be indebted to my supervisors Professor Des Higham and Professor Kerem Akartunalı from whom I have learned so much over the years. This thesis has been a work in progress for longer than I care to remember, and I am especially grateful for their enduring support and encouragement. I have no doubt that they may even be celebrating more than myself when I can firmly close this chapter of my life.

I would like to extend my thanks and appreciation to staff and fellow PhD students in the Department of Mathematics & Statistics for all their help throughout my time there. Strathclyde is a special place which will forever be held fondly in my memory as the place I met, my fiancée, Lyndsay.

Finally, I would like to thank my family. In particular, my heartfelt thanks go to my parents Sharon and William, my grandparents Barbara and James and my recently deceased grandparents Helen and David for all their love, encouragement and calming influence throughout my studies.

# Abstract

We derive new, exact expressions for network centrality vectors associated with classical Watts–Strogatz style "ring plus shortcut" networks. We also derive easy-to-interpret approximations that are highly accurate in the large network limit. The analysis helps us to understand the role of the Katz parameter and the PageRank parameter, to compare linear system and eigenvalue based centrality measures, and to predict the behavior of centrality measures on more complicated networks. We also derive accurate upper and lower bounds for the domiannt, Perron-Frobenius, eigenvalue of a "ring plus shortcut" network. The results are illustrated with computational experiments, and directions for future work are discussed.

# Contents

# Chapter 1

# Introduction

## 1.1 Outline of thesis

Algorithms that quantify the importance of nodes in a network are proving to be extremely useful [21, 41, 49]. They allow us to understand hierarchies, discover critical components, and identify targets for deeper investigation. Many of the key ideas behind network centrality measures arose out of the social sciences, where researchers were interested in understanding structural attributes of human interaction networks [23]. The ability to determine who or what is important is also valuable in many application areas, including healthcare, security, advertising, publishing and politics [3, 26, 35, 38].

A key issue, and perhaps a reason for the continued development of new ideas in the field, is that there is no universally-agreed definition (or set of definitions) for importance, and hence no gold-standards for judging centrality measures. So, issues such as validating implementations, understanding the role of algorithm parameters, and comparing centrality measures can only be partially addressed, typically by using real world data sets where some proxy for importance is available, leading to conclusions that are (a) empirically based and (b) problem-set dependent.

In this work, we contribute to the field by showing that there is a synthetic but widely studied "small world" type network for which we can analytically characterize and compare four well-known centrality measures—degree, Katz, eigenvector and PageRank. In particular, we can quantify how the node centralities change as the Katz parameter moves from zero (degree centrality) to its upper limit (eigenvector centrality). Moreover, we show how the same techniques allow us to characterize fully these measures on more complex networks where the performance can depend strongly on the choice of Katz or PageRank parameter.

Chapter 2 provides a brief background to the area of network science. Chapter 3 gives a brief overview of centrality measures and details Katz, eigenvector and PageRank centrality that we study in more detail in this thesis. Chapter 4 summarizes relevant work on Watts-Strogatz style small world networks and then we analyse the Katz and eigenvector centrality in the context of a Watts-Strogatz 'ring plus shortcut' setting. This 'ring plus shortcut' model analysis is extended to $M$-shortcuts across the ring and to $2n$-nearest neighbours in Chapter 5. In Chapter 6 the same style of analysis is applied to more general networks, giving Katz parameter cut-offs where the ranking from out-degree gives way to that of eigenvector centrality. PageRank centrality is analysed separately in Chapter 7 where the exact solution is found to have a more complicated non-monotonic form and we then analyse an example network where we find PageRank parameter cut-offs where the ranking of nodes change. We finish with a discussion in Chapter 8.

## 1.2 Publications and presentations

Some of the content presented within this thesis appears in the article

- M. Paton, K. Akartunalı and D.J. Higham, *Centrality analysis for modified lattices*, SIAM J. Matrix Anal. Appl., 38(3) (2017), 1055-1073. [45]

Results derived from this were presented at

- 27th Biennal Numerical Analysis Conference (2017), University of Strathclyde;
- Capita PhD Symposium (2016), University of Strathclyde.

# Chapter 2

# Background and notation

Within this Chapter we give a brief background to the area of network science and discuss some random networks that are pertinent to the work presented in later Chapters. We also provide some information on the notation and definition of terms used throughout this thesis.

## 2.1 Network science

The field of network science has experienced a renewed period of interest in the last few decades due in part to its applicability to modeling complex systems that surround us in our everyday lives [3, 26, 35, 38]. Indeed, as computing performance continues to increase and with an increasing amount of data around us the field continues to undergo an intense period of research activity. Applications to areas of healthcare, security, advertising, publishing and politics have demonstrated the area can be useful for extracting meaningful insights and predictions that may not be entirely obvious.

Network science is the branch of mathematics concerned with the modelling of complex systems - systems defined by entities and the pairwise interactions that exist between them. In short, network science is the study of complex systems such as biological, social, financial and transportation networks. It is by modelling the system as a mathematical network we can utilise ideas from other fields of mathematics such as graph theory, statistics and numerical linear algebra to gain further insight into the system. Its roots are in the area of *graph theory* with what is thought to be the first publication in the field by mathematician Leonhard Euler in 1736 [22]. The posed problem in the original publication has now become a staple example in almost any undergraduate mathematics class - the Bridges of Königsberg as depicted in Figure 2.1.

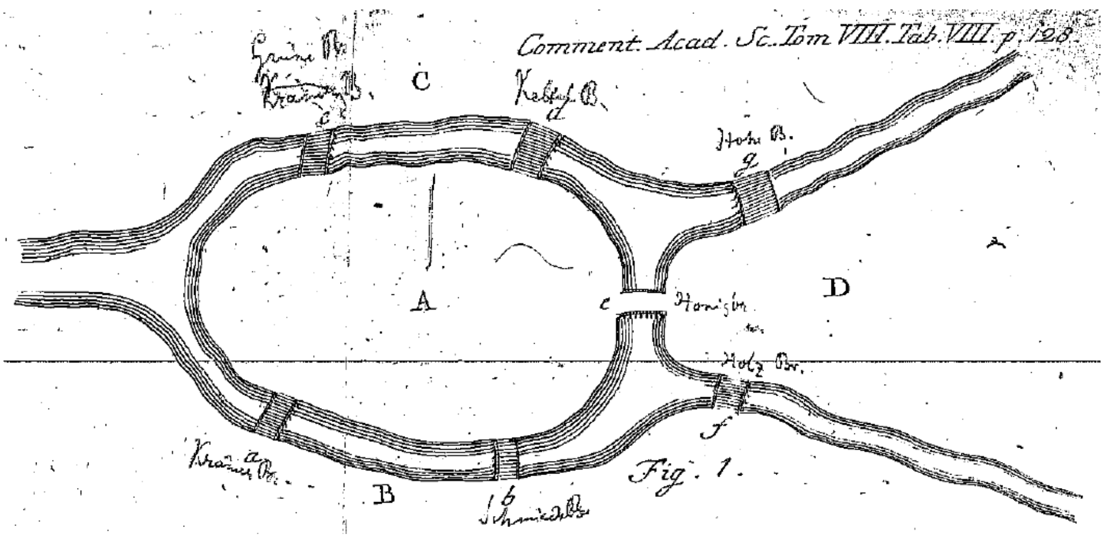



Figure 2.1: Bridges of Königsberg as depicted in [22]

The problem devised was to find a route through the city of Königsberg (Kingdom of Prussia now known as Kaliningrad, Russia) which was divided into four parts, two mainland portions and two islands, by the Pregal River such that each bridge was only used once. Euler shown that the problem could be posed in terms of the distinct land areas and the bridges which connect them, demonstrating that the absence of an *Eulerian path* rendered no solution. It was this paper by Euler that laid the foundations for what is now known as graph theory and introduced the idea of topology.

The study of random networks followed in the 20th century by Paul Erdős and Alfréd Rényi. Erdős and Rényi considered graphs with both a fixed number of nodes and edges with each possible graph equally likely to arise [20]. For example, if we consider the set of simple undirected graphs with 3 nodes and 2 edges there are a total of 3 defined graphs and each one may be induced with probability 1/3.

Edgar Gilbert later described a similar class of random networks defined by a fixed number of nodes but unlike that posed by Erdős and Rényi, Gilbert considered each edge to have an equal independent probability of existence [24]. Although these models are widely used, the consideration of each edge being equally likely to be present is not a good reflection of networks encountered in practice. In this sense, the Erdős–Rényi and Gilbert models are unlikely to capture the observed properties of many real-life networks.

Many networks exhibit properties that can be termed into the class of being 'small world'. These networks are typically classified by structures in which each node does not neighbour each other but the notion that neighbours of one node are common to neighbours of another. This implies that the average distance between any nodes (the number of edges traversed to get from node to node) is proportional to the logarithm of the number of nodes. A typical example of this would be in considering social friendships in society in which friendship circles overlap so although $A$ may not be friends with $C$ they may share a common friend $B$ such that a path from $A$ to $C$ in the network exists, thereby the cliché "it's a small world".

In the 1960s, social scientist Stanley Milgram noted the importance of the small world problem and its range of application areas being of interest to not just social scientists. In a real world experiment he set out to answer the question of how many intermediate people are needed before a route between any two people in the world is achieved [39]. The experiment was simple: person $A$ has to get a folder delivered to person $Z$ only by sending the folder to a person they personally knew on a first name basis. Each intermediary was to note their details in the folder which meant that an endless loop of passing the folder back to another that had already received it could be avoided. This stage of the process also allowed for further insight by studying the chains of transmission.

Milgram's experiment reported chains of transmission ranging from 2 to 10 with a reported median of 5. An important limitation of this experiment, that Milgram recognised, is that whilst each person in the chain may try to be most efficient in deciding who to pass the folder to it's inconceivable to think that the folder always followed the shortest possible path from person $A$ to person $Z$. Analysis of the transmissions highlighted that the participants were roughly three times more likely to send the folder on to someone of the same sex and almost six times more likely to send to a friend as opposed to a relative.

As the need to model and understand real-life phenomena grew it became more important to develop classes of networks that mirrored observed properties. Seminal work of Watts and Strogatz published in 1998 did just that by defining a class of random networks that mirrored properties of real-life networks [50]. Their idea was to start with a regular $k$-neighbour ring and rewire each edge with independent probability $p$. Here, rewiring means replacing the "target" node by a node chosen uniformly at random. As the probability of rewiring each edge increases from zero to one we interpolate between the original $k$-neighbour ring and a completely random graph as depicted in Figure 2.2. It is within this range of probability that we observe a regime with lattice-like structure where the shortcuts allow efficient traversal around the network.

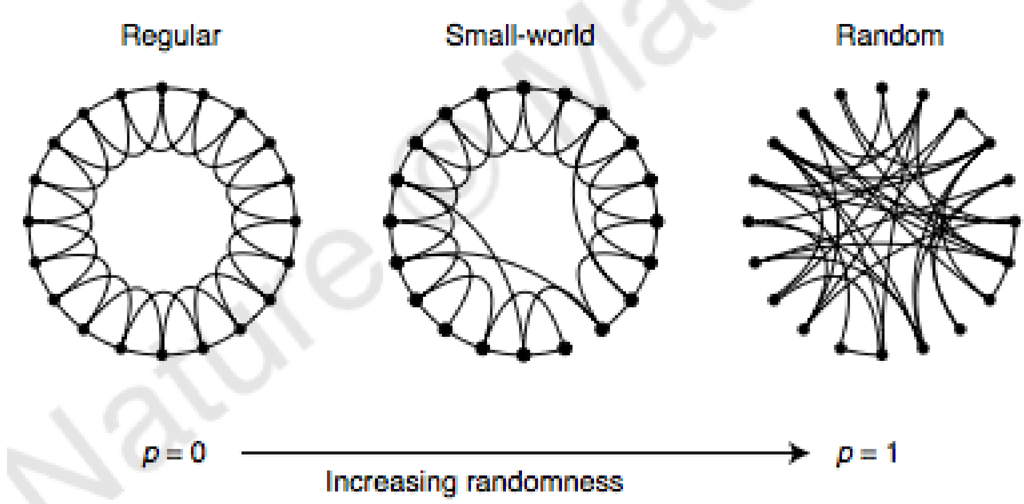



Figure 2.2: Small world network as depicted in [50]

The work of Watts and Strogatz focused on two independent properties – shortest path length connecting nodes and the clustering coefficient – as randomness is added to the regular lattice. Watts and Strogatz defined a node clustering coefficient as the proportion of edges each node has compared to a complete graph in which all nodes are connected and subsequently averaged over all nodes to form a single network measure. They found there was an interval for the choice of rewiring probability $p$ that give a shortest path length almost equivalent to a random graph but retained a significantly higher clustering coefficient than a random graph. Watts and Strogatz noted that many real-life world networks have higher clustering coefficient than what should be expected of a random graph. It is the clustering coefficient property of their proposed class of networks that distinguishes it from previous work of Erdős–Rényi and Gilbert.

## 2.2 Notation & definitions

In its most basic form, a network is simply a collection of nodes and edges. Depending on the application area, a node can represent an object, a person or even a place while an edge would illustrate a connection or some form of pairwise interaction.

A set is a well-defined collection of objects and we refer to the objects that belong to a set as being elements or members of that set. A rule of thumb is we represent sets by upper case letters and its elements by lower case letters. We write $a \in A$ if $a$ is an element of the set $A$ and $a \notin A$ if $a$ is not an element of the set $A$. A set is defined by some condition or by giving an explicit list of its members, for example $A = \{1, 2, 3\}$ or $B = \{b^3 \text{ such that } 0 < b < 1\}$.

A network or graph is defined by two sets - one set of nodes and one set of edges. Typically, the nodal set will be denoted by the set $V$ (for vertices) and the edge set by $E$. An element of an edge set is an ordered pair $(i, j)$, where $i, j \in V$, meaning an edge originates from node $i$ and terminates at node $j$. The graph induced by the set of nodes $V$ and the set of edges $E$ is given by $G(V, E)$. Each node in the network should be uniquely identifiable.

For example, in Figure 2.3 there are a total of 5 nodes. As each edge is undirected (indicated by lack of direction on edges) the edge set must contain 14 elements.

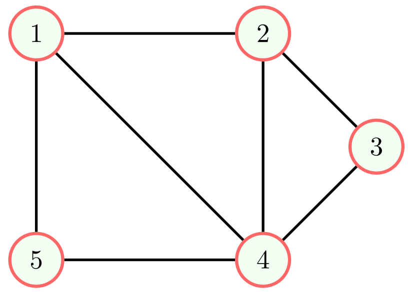



Figure 2.3: Example network with 5 nodes and 7 edges.

This graph is induced by the nodal and edge set

$$\begin{aligned} V &= \{1,2,3,4,5\} \\ E &= \{(1,2),(1,4),(1,5),(2,1),(2,3),(2,4),(3,2),(3,4), \\ &\quad (4,1),(4,2),(4,3),(4,5),(5,1),(5,4)\}. \end{aligned}$$

Note how $|V| = 5$ and $|E| = 14$ meaning there are 5 nodes and 14 directed edges (7 undirected edges). We will denote the number of nodes in a network by $N$.

A traversal of edges around the network

$$(v_1, v_2), (v_2, v_3), \ldots, (v_{p-1}, v_p)$$

is called a *walk*. The length of a walk is the number of edges traversed. A *trail* is a walk in which the edges traversed are unique. If the nodes in the walk are unique then it follows that the edges are also unique and this is called a *path*.

There are many well established ways to represent a network mathematically such as an edge or adjacency list or an adjacency matrix. Each representation has its own advantages and disadvantages, for example, in the case of a large sparse network it would be much more efficient to only store the sets of edges which had a connection rather than store a full matrix. Due to the reliance on linear algebra for many applications in network science it is common to work with the adjacency matrix which is typically sparse.

A network's adjacency matrix, $A \in \mathbb{R}^{N \times N}$, is a square binary matrix where a row/column represents an element in the nodal set $V$ of cardinality $N$. If

$A = (a_{ij})_{1 \leq i,j \leq N}$ then

$$a_{ij} = \begin{cases} 1, & \text{if } (i,j) \in E, \\ 0, & \text{otherwise.} \end{cases}$$

That is, we put a 1 in position $i, j$ if there is an edge connecting node $i$ to node $j$ and put a 0 if there is no such edge. A network is said to be *simple* if there is no self–loop in the network (i.e., $a_{ii} = 0$ for all $i$), no multiple edges between nodes and the edges are undirected meaning information can travel in both directions. In this case the diagonal elements of the adjacency matrix are zero and $A$ is symmetric. Our work will primarily be focused on unweighted directed networks without self–loops, however the adjacency matrix definition can be extended to the case of weighted networks with $a_{ij} = w_{ij}$ if required.

As the adjacency matrix $A$ details the connections between nodes in the network it can be used to compute walks around the network with $A^k$ being a count of walks of length $k$ between each pair of nodes.

The spectral radius of the adjacency matrix $A$ is defined as

$$\rho(A) = \max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_N|\},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_N$ are the eigenvalues of $A$ such that $Ay_i = \lambda_i y_i$.

Throughout this thesis we let $\mathbf{1} \in \mathbb{R}^N$ denote the vector of ones and $A^T$ be the transpose of $A$ such that $(A^T)_{ij} = (A)_{ji}$.

In this work we encounter palindromic polynomials of even degree with real coefficients and therefore we state here some useful properties of these polynomials that we rely upon later. Given a polynomial

$$p(z) = a_0 + a_1 z + \ldots + a_n z^n = \sum_{i=0}^{n} a_i z^i,$$

the reciprocal polynomial (a reversal in coefficients) is given by

$$p^*(z) = a_n + a_{n-1} z + \ldots + a_0 z^n = \sum_{i=0}^{n} a_{n-i} z^i.$$

A polynomial is self-reciprocal if $p(z) = p^*(z)$ for all $z$, and therefore $a_i = a_{n-i}$ [46]. In this case the coefficients form a palindrome and we use the term palindromic polynomial. A well known palindromic polynomial is $(z+1)^n$ where the coefficients are given by the $n$-th row of Pascal's triangle.

The reciprocal polynomial can be written in terms of the original polynomial as $p^*(z) = z^n p(z^{-1})$. It follows that in the case of a palindromic polynomial we have $p(z) = z^n p(z^{-1})$. Therefore if $z_1 \neq 0$ is a root of a palindromic polynomial then $z_1^{-1}$ is also a root of the palindromic polynomial. That is, roots of palindromic polynomials appear in reciprocal pairs.

If we have a polynomial with real coefficients it follows that if $z_1 \in \mathbb{C}$ is a root then the complex conjugate, $\bar{z_1} \in \mathbb{C}$, is also a root [31].

In the context of the palindromic polynomials, if we have a complex root then the inverse of the complex root is also a root and so is the complex conjugate of both, i.e., $z_1, z_1^{-1}, \bar{z_1}, \bar{z_1}^{-1}$ are all roots.

We note here that if a complex root lies on the unit circle, i.e., has modulus one, then the inverse is equal to the conjugate. In this special instance this gives rise to two distinct roots rather than four.

# Chapter 3

# Centrality measures

Given a network adjacency matrix, $A$, a centrality measure assigns a value $x_i > 0$ to each node $i$, with a larger value indicating a greater level of importance. Typically, it is the ranking of the centrality values that matters—we only care whether one node is more or less important than another, so the vector $x \in \mathbb{R}^N$ is equivalent to $\beta x + \gamma \mathbf{1}$ for any $\beta, \gamma > 0$.

Within this chapter we explore a brief history of centrality measures and introduce Katz, eigenvector and PageRank centralities that our work is built upon.

## 3.1 History

Many of the key ideas behind network centrality measures arose in the social sciences, where researchers wanted to quantify the importance of some individuals in human-interaction networks by studying the structural attributes of the network [23].

In 1948 Alex Bavelas first hypothesised a link between centrality and influence in communication networks [5]. Subsequent research, led by Bavelas, at Massachusetts Institute of Technology (MIT) in the late 1940s and early 1950s involved a problem solving exercise. Participants were given different information and the group had to piece all the information together before taking decisions and solving the problem. The research concluded there was a relationship between leadership, efficiency and organisation of the group and centrality in communication patterns [6, 7, 36].

This research carried out at MIT motivated others to get involved and experiments of similar and more complicated guises followed whereby different communication strategies could be employed to see the impact on efficiency at problem solving. Robert Burgess provides a summary of the subsequent experiments but notes that the "research has not provided consistent and cumulative results" [15]. It is clear that centrality is related to the influence of outcomes, for example, if we were to examine a company's email communications we would be able to identify those that are in charge as information typically flows down a hierarchy.

The early idea of centrality was also being used in other contexts throughout the 1960s and 1970s. For example, researchers have used centrality to understand political integration in India [18] and diffusion in the steel industry [19].

Freeman explains that whilst the idea of centrality is accepted and in use in a wide variety of applications there is no consensus on what exactly centrality is or indeed what exactly it should measure [23]. Therefore there exists a plethora of centrality measures in use today, each of which has its own definition of importance.

Although the term centrality was coined in the late 1940s the fundamental ideas and notions that lay behind it have a longer history dating back to the latter part of the 19th century in the context of ranking chess players. In a round robin competition players would amass points – 1 for a win; 1/2 for a draw; 0 for a loss. In today's terminology this would form the basis for a weighted adjacency matrix.

Whilst it would perhaps be reasonable to assume that a player's overall number of points should determine the rankings of those in the competition, Oscar Gelbfuhs and later Hermann Neustadl argued that this however gives equal weight to a win against a weak or a strong player and suggested that points should be weighted according to the strength of the opposition [1].

Whilst there is merit in the argument made by Gelbfuhs and Neustadl it proved to be controversial as it could give rise to determining a tournament winner not because they are the strongest player but rather their opponents were strong. Outraged at the suggestion, a young Edmund Landau got involved and instead proposed that the overall rankings should be based on an eigenvector, of the now known, adjacency matrix [33]. Interestingly, one of Landau's points of criticism was that such a method proposed by Gelbfuhs and Neustadl would produce a different result on every iteration even though, as we now know through analysis of the power method, the iteration converges to the aforementioned eigenvector. The work of Landau in terms of ranking methods in a range of different contexts is discussed in [9].

This notion is widely accepted today and is fundamental to what is perhaps the most commonly used ranking algorithm – Google's PageRank [44]. Used by millions on a daily basis to find and rank web pages according to the search criteria the PageRank centrality is based on the idea that an important web page is one to which many other important web pages link.

Whilst there are lots of centrality measures with different proxies for what it means to be important, the rise of digital information and online social media in a range of applications such as advertising, defence and security has meant that what are known as spectral methods have become more important than the classical alternatives. As with the examples touched upon above, spectral methods have the property that a measure of importance and hence ranking is based upon mutual reinforcement, i.e., a good chess player is someone that wins against other good chess players and an important web page is one that other important web pages link to. The history of spectral ranking is detailed in [48].

## 3.2 Spectral centrality measures

We summarize here the concepts of Katz, eigenvector and degree centrality, referring to [2, 11, 21, 41, 49] for historical details and discussions of implementation issues. We also consider the PageRank algorithm [25, 34, 44] which has a different feel; summarizing incoming, rather than outgoing, information, but also assigning a positive real value to each node. Our overall aim is to study, in a specific setting, how changes to the network affect centrality. We note that related questions have been addressed in other contexts; see, for example, [10, 17].

A network is strongly connected if there exists a path between every pair of nodes in the network. That is, there exists a sequence of edges that allows every node to be reached from every other node. We now assume that the networks are strongly connected so that the adjacency matrix, $A$, is irreducible. This property holds for the network class that we study in subsequent chapters.

Katz centrality [32] defines $x$ via the linear system

$$(I - \alpha A)x = \mathbf{1}, \tag{3.1}$$

where $\alpha \in (0, 1/\rho(A))$ is a free parameter. Several authors have suggested particular choices for $\alpha$; see [2] and the references therein. This measure can be motivated by expanding the resolvent $(I - \alpha A)^{-1}$ to give

$$x = \left(I + \alpha A + \alpha^2 A^2 + \alpha^3 A^3 + \cdots\right) \mathbf{1}.$$

Noting that $(A^k)_{ij}$ counts the number of distinct walks of length $k$ from $i$ to $j$, we see that $x_i$ is a weighted sum of the number of walks from node $i$ to all other nodes, where the count for walks of length $k$ is scaled by $\alpha^k$. So $x_i$ is a measure of how well node $i$ can send information around the network, with more weight given to shorter traversals.

In the limit as $\alpha \to 0$ from above the ranking given by (3.1) coincides with the ranking from degree centrality, which arises from taking $x_i$ to be the out-degree (number of edges originating from the node) of node $i$; that is,

$$x_i = \text{outdeg}_i, \quad \text{where outdeg} = A\mathbf{1}. \tag{3.2}$$

This makes sense intuitively, since accounting only for the shortest possible walks—of length one—is equivalent to computing the out-degree.

Eigenvector centrality can be motivated recursively, with a node being important if it has links to important nodes. This leads to the expression

$$x_i \propto \sum_{j=1}^{N} a_{ij} x_j. \tag{3.3}$$

Letting $1/\lambda$ denote the constant of proportionality, we may write

$$Ax = \lambda x, \tag{3.4}$$

showing that $x$ must be an eigenvector of $A$ corresponding to eigenvalue $\lambda$. Requiring $x_i > 0$ for all $i$ forces $\lambda$ and $x$ to be the Perron-Frobenius eigenvalue and eigenvector, respectively. For a discussion of Perron-Frobenius theory, we refer to [30]. Here we note that $\lambda = \rho(A)$ is the dominant eigenvalue of $A$, under our assumption that $A$ is irreducible.

In the limit as $\alpha$ tends to $1/\rho(A)$ from below, the Katz centrality vector in (3.1) gives the same results as the eigenvector centrality vector; see, for example, [8].

The PageRank algorithm, originally proposed in [44], can be motivated from several different perspectives. For example, keeping in mind the context of web pages, we could alter (3.3) by arguing that the importance of a node is dependent on the importance of the nodes that point to it. If $a_{ij} = 1$ indicates a hyperlink *from* node $i$ *to* node $j$, then we can set up an iteration where each node is given a convex combination of a basal score and a normalized sum of the scores of its followers, so that

$$x_i^{[k+1]} = 1 - d + d \sum_{j=1}^{N} \frac{a_{ji}}{\mathrm{outdeg}_j} x_j^{[k]}.$$

Here, scaling by $\mathrm{outdeg}_j$ ensures that every node has the same opportunity to distribute its influence across its neighbours. The parameter $0 < d < 1$ controls the relative weight given to this redistribution. Taking the limit $k \to \infty$, we define the PageRank vector as the solution of

$$\left(I - dA^T D^{-1}\right) x = (1 - d)\mathbf{1}, \tag{3.5}$$

where $D = \mathrm{diag}(\mathrm{outdeg}_j)$. We refer to [25, 34] for further details.

# Chapter 4

# Analysis of modified rings

Chapter 4 introduces the network class, a nearest neighbour periodic ring with a shortcut, that we study centrality measures on. This chapter focuses on Katz and eigenvector centality demonstrating a similar form of solution to both but one central difference – Katz centrality tends to a fixed non-zero value whilst eigenvector tends to zero on moving away from the spike node. We also include bounds on the Perron-Frobenius eigenvalue associated with the adjacency matrix that justifies a restriction of the Katz parameter domain in our analysis.

## 4.1 Small world models

As mentioned in Chapter 1, the seminal work of Watts and Strogatz [50] introduced a class of random graphs characterised by having many “local” links and a few “long range” links. Those authors showed, via computational experiments, that such a model can reproduce clustering and pathlength properties that have been observed in real world complex networks. A key idea in [50] was to add randomness to a regular lattice. Starting from an undirected periodic ring with fixed-range nearest-neighbour connections, the authors introduced a rewiring procedure—each node in the lattice was examined in turn, and, for each of its undirected links, with small independent probability the end point of that link was replaced by another node chosen uniformly at random. In terms of developing analytical results to back up the observations in [50], the rewiring process presents difficulties; for example, a node may become isolated with nonzero probability. For this reason, subsequent research focused on a slight variation where existing edges are not altered, but new edges, termed *shortcuts*, are added between randomly selected nodes in the network. For example, focusing on the $N \to \infty$ limit, [40, 42, 43] develop heuristic approximations and [4] gives rigorous results. Markov chain versions with hitting time (mean number of edges traversed to get from node $i$ to node $j$ on a random walk for the first time) as a proxy for pathlength were also studied rigorously in [16, 28, 29, 51]. Of particular relevance to our study is the reference [37], which analysed the effect of shortcuts on the underlying matrix spectrum from a linear algebra viewpoint.

In this work we will therefore use the shortcut concept, rather than rewiring. The details are described in the next section. We also note that in the case where each undirected edge in the underlying periodic ring is regarded as an independent pair of directed edges, so that a rewired edge produces a directed link, the study of Katz and eigenvalue centralities is not interesting. This can be seen from the following simple lemma.

**Lemma 4.1.** *Suppose that all nodes in a network have the same out-degree. Then all nodes have the same Katz centrality measure. Similarly, assuming strong connectivity, all nodes have the same eigenvector centrality measure.*

*Proof.* We have

$$A\mathbf{1} = \mathrm{od}\,\mathbf{1}, \tag{4.1}$$

where $\mathrm{od} \equiv \mathrm{outdeg}_i$ denotes the common out-degree. Hence $\mathbf{1}$ is the Perron-Frobenius eigenvector and $\rho(A) = \mathrm{od}$. Any Katz parameter $0 < \alpha < 1/\mathrm{od}$ is valid, and we see from (4.1) that $x = \mathbf{1}/(1 - \alpha\,\mathrm{od})$ solves the Katz system (3.1). □

## 4.2 Matrix modification

Let $C \in \mathbb{R}^{N\times N}$ be the adjacency matrix for the periodic nearest neighbour ring; so in the $N = 6$ case,

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \tag{4.2}$$

We note that $C$ is symmetric and circulant, and standard theory shows that the Perron-Frobenius eigenvalue is $\rho(C) = 2$ with corresponding eigenvector $x = \mathbf{1}$ [30, page 100], which, from Lemma 4.1, also solves the Katz system (3.1), up to a multiplicative factor, for any $0 < \alpha < 1/\rho(C)$. It is, of course, intuitively reasonable that all nodes in this network should be assigned the same centrality value in (3.1), (3.2) and (3.4).

Next we add a single directed shortcut. Without loss of generality we give the shortcut to node 1 and let $L$ be the index of the target node. So our adjacency matrix $A$ in (3.1) has the form $A = C + E$, where the rank one matrix $E$ is zero except for $E(1, L) = 1$. Liu, Strang and Ott [37] describe this as a *modification* of $C$, to emphasize that we have an $O(1)$ change in a matrix entry, rather than the type of small change studied in classical matrix perturbation theory. These authors studied the eigenvector associated with the dominant eigenvalue of $A$, and related matrices, and constructed accurate approximations to this vector. Further work concerning the eigenvalues arising from general modifications to structured matrices has appeared in, for example, [12, 13, 14].

Our work is strongly motivated by [37] but differs from it in three respects.

- Rather than deriving small residual approximations and then using stability arguments to bound the forward error, we construct exact solutions that can be expanded asymptotically. This more direct route leads to shorter proofs and sharper bounds.

- We consider Katz and PageRank centrality (as well as the eigenvalue problem).

- We interpret the results from a network science perspective and show how they can give new insights about behaviour on more complicated networks.

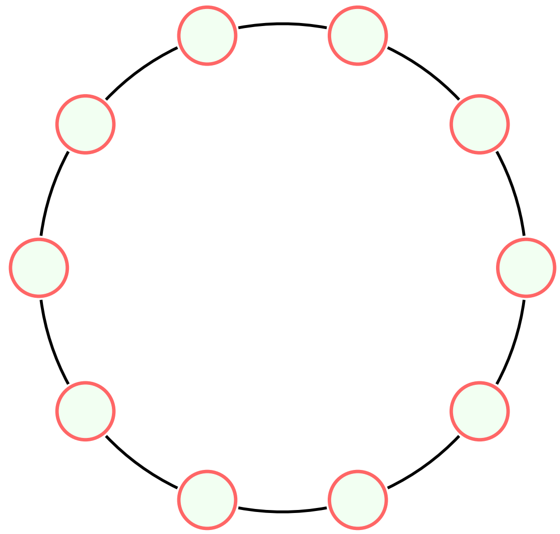
For convenience, we let $p(i)$ for any $1 \leq i \leq N$ denote the periodic distance from node $i$ to node 1, that is,

$$p(i) = \min\left(|i-1|, |N-(i-1)|\right). \tag{4.3}$$

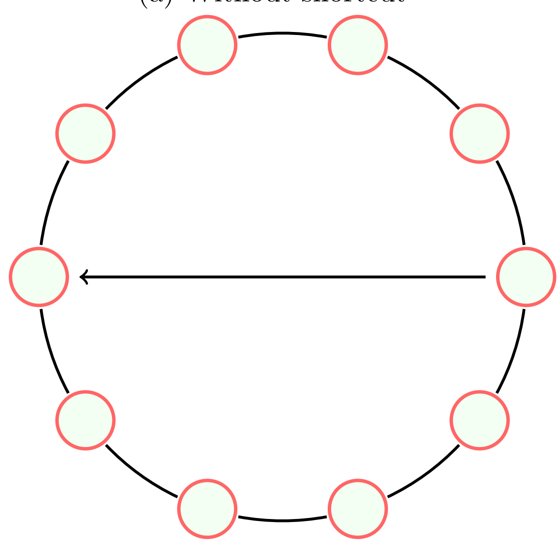
We can assume without loss of generality that the receiving node $L$ is not beyond the half way, or "six o'clock", position on the ring. We are interested in large networks with long-range shortcuts. So, letting $\lfloor \cdot \rfloor$ denote the integer part, for some fixed proportion $0 < \theta \leq 1$ we set

$$L = \begin{cases} \lfloor \theta(N/2+1) \rfloor & \text{when } N \text{ is even,} \\ \lfloor \theta(N+1)/2 \rfloor & \text{when } N \text{ is odd.} \end{cases} \tag{4.4}$$

We note that $L \to \infty$ as $N \to \infty$.



(a) Without shortcut



(b) With shortcut

Figure 4.1: Illustration of nearest neighbour periodic ring with $N = 10$

## 4.3 Exact solution for Katz centrality

In the upper picture of Figure 4.2, the asterisks show Katz centrality values for a network with adjacency matrix $A = C + E$, that is, components of $x$ from (3.1). We chose a small network size in order to make the key effects visible. More precisely, we used an $N = 20$ node ring with a shortcut from node 1 to node $L = 8$, for $\alpha = 0.3$. Because node 1 owns the extra, long-range edge, it attains the highest centrality score, at around 3.5. The most distant node, periodically, node 11, is deemed the least central. Insight from [37], or from eyeballing the solution, suggests that components of $x_i$, when suitably shifted, might be varying geometrically as the index $i$ moves periodically around the ring; that is $x_i = b + ht^{p(i)}$ for some constants $b, h > 0$ and $0 < t < 1$. Fitting an ansatz of this form leads us to the circles in the upper picture of Figure 4.2. The agreement is close—below $2 \times 10^{-5}$ in Euclidean norm.

The lower picture in Figure 4.2 shows, on a log scale, the discrepancy between those asterisks (true solution) and circles (geometric decay ansatz). We see a very small contribution that, in contrast to the overall solution, *grows* geometrically as we move periodically away from node 1.

We now state a theorem that explains Figure 4.2. The theorem concerns the limit $N \to \infty$ with a fixed Katz parameter $0 < \alpha < 1/2$. This upper limit for $\alpha$ is chosen because, as proved in Section 4.5, $\rho(A)$ tends to 2 from above as $N \to \infty$.

**Theorem 4.1.** *For the undirected ring plus directed shortcut network with adjacency matrix $A = C + E$, for sufficiently large $N$ the unique solution of the Katz system (3.1) has the form*

$$x_i = b + h_1 t_1^{p(i)} + h_2 t_2^{p(i)}. \tag{4.5}$$

*Here, $b, t_1, t_2, h_1, h_2$ are constants, i.e., independent of $i$, and $b, t_1, t_2$ are also independent of $N$. In particular, $h_1 > 0$, $h_2 > 0$, $b = 1/(1 - 2\alpha)$ and $t_1, t_2$ are the roots of the palindromic quadratic $\alpha t^2 - t + \alpha$, so that*

$$t_1 = \frac{1 - \sqrt{1 - 4\alpha^2}}{2\alpha}, \quad t_2 = \frac{1 + \sqrt{1 - 4\alpha^2}}{2\alpha}, \tag{4.6}$$

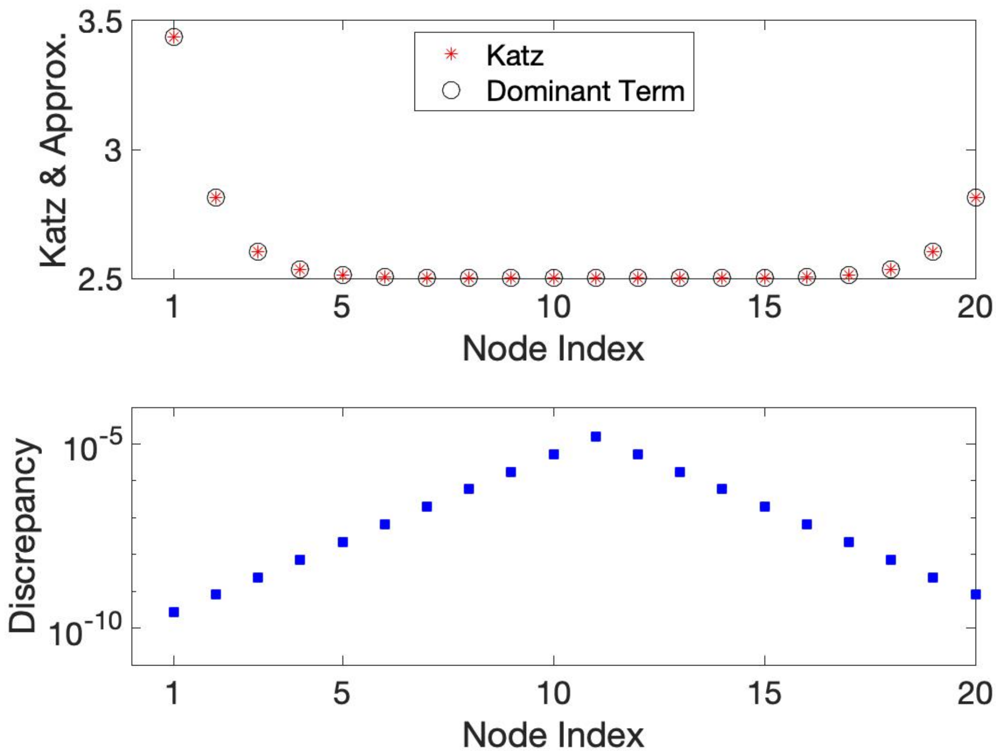



Figure 4.2: Upper picture: asterisks show components of Katz vector from (3.1) and circles show the approximation $b + h_1 t_1^{p(i)}$ from (4.7). Here $b = 1/(1 - 2\alpha)$, $t_1$ is defined in (4.6) and $h_1$ was found by solving (4.15). Lower picture: the discrepancy $x_i - b - h_1 t_1^{p(i)}$ on a log scale. From Theorem 4.1, this quantity has the form $h_2 t_2^{p(i)}$, and hence grows geometrically away from the shortcut node. However, it is uniformly $O(t_1^{N/2})$ for a fixed $0 < t_1 < 1$, and hence rapidly becomes negligible as the network size $N$ increases.

*with* $t_2 = 1/t_1$ *and* $0 < t_1 < 1 < t_2$. *Moreover, the final term in (4.5) is exponentially small asymptotically, in the sense that for all* $1 \le i \le N$,

$$x_i = b + h_1 t_1^{p(i)} + O(t_1^{N/2}), \tag{4.7}$$

*with* $h_1 = O(1)$.

*Proof.* We begin with the ansatz

$$x_i = b + h t^{p(i)}. \tag{4.8}$$

The Katz system (3.1) essentially reduces to three scalar equations where we need to consider:

- General node $k$ where each neighbouring node, $k-1$ and $k+1$, differs by a factor of $t$
- Node 1 which has a shortcut across the ring to node $L$
- The node(s) furthest away from node 1 periodically, i.e.
  - node $q = N/2 + 1$, if $N$ is even
  - node $q = (N+1)/2$ or $q = (N+3)/2$, if $N$ is odd.

At a general node $k$, corresponding to the $k$th row of the linear system, for $k \neq 1$ and $k \neq q$ we have

$$-\alpha x_{k-1} + x_k - \alpha x_{k+1} = 1. \tag{4.9}$$

Inserting (4.8), this becomes

$$b(1-2\alpha) + h t^{p(k)} \left[1 - \alpha t - \alpha/t\right] = 1.$$

We can satisfy this equation independently of $k$ by setting $b = 1/(1-2\alpha)$ and choosing $t$ to be either root of the quadratic $\alpha t^2 - t + \alpha$. By linearity of (4.9), we may extend (4.8) to include a linear combination involving both roots, so our ansatz becomes

$$x_k = b + h_1 t_1^{p(k)} + h_2 t_2^{p(k)}, \tag{4.10}$$

where

$$t_1 = \frac{1-\sqrt{1-4\alpha^2}}{2\alpha}, \quad t_2 = \frac{1+\sqrt{1-4\alpha^2}}{2\alpha}.$$

Because the quadratic is palindromic, $t_1$ and $t_2$ must satisfy $t_1 t_2 = 1$. It is clear that these roots are real, with $0 < t_1 < 1 < t_2$, and we note the further useful facts

$$1 - 2\alpha t_1 = \sqrt{1-4\alpha^2} > 0, \qquad 1 - 2\alpha t_2 = -\sqrt{1-4\alpha^2} < 0 \tag{4.11}$$

and

$$(1-\alpha)t_1 - \alpha = \alpha t_1(t_1 - 1) < 0, \qquad (1-\alpha)t_2 - \alpha = \alpha t_2(t_2 - 1) > 0, \tag{4.12}$$

The two remaining equations to satisfy from (3.1) arise at nodes 1 and $q$, where we require

$$-\alpha x_N + x_1 - \alpha x_2 - \alpha x_L = 1 \tag{4.13}$$

and

$$\begin{cases} -\alpha x_{N/2} + x_{N/2+1} - \alpha x_{N/2+2} = 1, \text{ if } N \text{ is even} & \text{(4.14a)} \\ -\alpha x_{(N-1)/2} + x_{(N+1)/2} - \alpha x_{(N+3)/2} = 1, \text{ if } N \text{ is odd} & \text{(4.14b)} \end{cases}$$

respectively. In (4.14b) above, we have used $q = (N+1)/2$ but $q = (N+3)/2$ is equally valid and will result in the same analysis.

**Consider the case where the number of nodes, $N$, is even.**

Inserting (4.10) into (4.13) and (4.14a), using $b(1-2\alpha) = 1$, we have the linear system

$$\begin{bmatrix} 1 - 2\alpha t_1 - \alpha t_1^{L-1} & 1 - 2\alpha t_2 - \alpha t_2^{L-1} \\ t_1^{N/2-1}(t_1 - 2\alpha) & t_2^{N/2-1}(t_2 - 2\alpha) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} b\alpha \\ 0 \end{bmatrix}. \tag{4.15}$$

To prove that a solution of the form (4.10) exists, we now show that this system is nonsingular. The determinant may be written as

$$\left(1 - 2\alpha t_1 - \alpha t_1^{L-1}\right) t_2^{N/2}(1 - 2\alpha t_1) - \left(1 - 2\alpha t_2 - \alpha t_2^{L-1}\right) t_1^{N/2}(1 - 2\alpha t_2). \tag{4.16}$$

Recalling that $0 < t_1 < 1 < t_2$, we see from (4.11) that, as $N \to \infty$ and hence $L \to \infty$, the first term in (4.16) becomes, to leading order,

$$(1 - 2\alpha t_1)^2 \, t_2^{N/2} \to \infty.$$

The second term in (4.16) becomes, to leading order,

$$\alpha t_2^{L-1} t_1^{N/2} (1 - 2\alpha t_2) = \alpha t_1^{N/2-L} (t_1 - 2\alpha). \tag{4.17}$$

Recall that $L = \lfloor \theta(N/2 + 1) \rfloor$, so

$$N/2 - L > N/2 - \theta(N/2 + 1) = (1 - \theta)N/2 - \theta.$$

Hence, as $N \to \infty$, the term $\alpha t_1^{N/2-L}(t_1 - 2\alpha)$ in (4.17) is bounded for $\theta = 1$, and tends to zero for $\theta < 1$. So the first term in (4.16) dominates, and the determinant is bounded away from zero for large $N$.

**Consider the case where the number of nodes, $N$, is odd.**

Inserting (4.10) into (4.13) and (4.14b), using $b(1 - 2\alpha) = 1$, we have the linear system

$$\begin{bmatrix} 1 - 2\alpha t_1 - \alpha t_1^{L-1} & 1 - 2\alpha t_2 - \alpha t_2^{L-1} \\ t_1^{(N-3)/2} \left[(1-\alpha)t_1 - \alpha\right] & t_2^{(N-3)/2} \left[(1-\alpha)t_2 - \alpha\right] \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} b\alpha \\ 0 \end{bmatrix}. \tag{4.18}$$

To prove that a solution of the form (4.10) exists, we now show that this system is nonsingular. The determinant may be written

$$\left(1 - 2\alpha t_1 - \alpha t_1^{L-1}\right) t_2^{(N-3)/2} \left[(1-\alpha)t_2 - \alpha\right] - \left(1 - 2\alpha t_2 - \alpha t_2^{L-1}\right) t_1^{(N-3)/2} \left[(1-\alpha)t_1 - \alpha\right]. \tag{4.19}$$

Recalling that $0 < t_1 < 1 < t_2$, we see from (4.11) and (4.12) that, as $N \to \infty$ and hence $L \to \infty$, the first term in (4.19) becomes, to leading order,

$$(1 - 2\alpha t_1) \, t_2^{(N-3)/2} \left[(1-\alpha)t_2 - \alpha\right] \to \infty.$$

The second term in (4.19) becomes, to leading order,

$$\alpha t_2^{L-1} t_1^{(N-3)/2}\left[(1-\alpha)t_1-\alpha\right] = \alpha t_1^{(N-1)/2-L}\left[(1-\alpha)t_1-\alpha\right]. \qquad (4.20)$$

Recall that $L = \lfloor \theta((N+1)/2) \rfloor$, so

$$(N-1)/2 - L > (N-1)/2 - \theta((N+1)/2) = (1-\theta)(N-1)/2 - \theta.$$

Hence, as $N \to \infty$, term $\alpha t_1^{(N-1)/2-L}\left[(1-\alpha)t_1-\alpha\right]$ in (4.20) is bounded for $\theta = 1$, and tends to zero for $\theta < 1$. So the first term in (4.19) dominates, and the determinant is bounded away from zero for large $N$.

**Continuation of general case: $N$ even or odd.**

We have shown that for sufficiently large $N$ the unique solution of the Katz system (3.1) has the form (4.10). We now follow up by showing that, in the exact solution (4.10), the growing term $h_2 t_2^{p(i)}$ is negligible for large $N$.

To do so, we first show that the second equations in (4.15) and (4.18) can both be reduced to the same condition. Using (4.15) and (4.11),

$$h_2 = -\frac{t_1^{N/2-1}(t_1-2\alpha)}{t_2^{N/2-1}(t_2-2\alpha)} h_1 = -t_1^N \frac{1-2\alpha t_2}{1-2\alpha t_1} h_1 = h_1 t_1^N. \qquad (4.21)$$

Similarly, using (4.18) and (4.12),

$$h_2 = -\frac{t_1^{(N-3)/2}\left[(1-\alpha)t_1-\alpha\right]}{t_2^{(N-3)/2}\left[(1-\alpha)t_2-\alpha\right]} h_1 = -t_1^{N-3}\frac{t_1^2(1-t_2)}{t_2(t_2-1)} h_1 = h_1 t_1^N. \qquad (4.21)$$

Hence, in the first equation of (4.15) and (4.18),

$$h_1\left[1 - 2\alpha t_1 - \alpha t_1^{L-1} - \alpha t_1^{N-L+1} - 2\alpha t_1^{N-1} + t_1^N\right] = b\alpha.$$

We see that $h_1 = O(1)$. Also, using (4.11), we deduce that $h_1 > 0$ for large $N$. It then follows from (4.21) that $h_2 > 0$ and $h_2 = O(t_1^N)$. Since the term

$h_2 t_2^{p(k)} = h_1 t_1^{N-p(k)}$ is largest when $k = N/2 + 1$, we also see that

$$x_i = b + h_1 t_1^{p(i)} + O(t_1^{N/2}).$$

$\square$

## 4.4 Exact solution for eigenvector centrality

This section looks at eigenvector centrality for the matrix $C + E$. We note that $\rho(C + E) \leq ||C + E||_\infty = 3$. Also, it is true in general that adding an edge to a strongly connected network strictly increases the spectral radius; see [30, Problem 8.4P14]. So $2 < \rho(C + E)$. From the network centrality perspective, we are concerned mainly with the structure of the Perron-Frobenius eigenvector. However, for completeness, in Section 4.5 we establish a tight bound on the corresponding eigenvalue.

The asterisks in the upper picture of Figure 4.3 show the components of the Perron-Frobenius vector for $A = C + E$ in the case where $N = 40$ and $L = 12$. As in the Katz case seen in Figure 4.2, there is evidence of periodic geometric decay. The circles in the picture show an ansatz of the form $x_i = s_1^{p(i)}$ for $0 < s_1 < 1$. Both vectors were normalized to have unit Euclidean norm. The lower picture shows the discrepancy between the two, and again we see a small contribution that increases periodically away from node 1.

Theorem 4.2 makes these observations concrete. The result is strongly motivated by the approach in [37], where exponentially accurate approximations were constructed.

In [37], Liu, Strang and Ott consider the case of a string of nodes which connect to each nearest neighbour. This can be represented as an adjacency matrix, $A \in \mathbb{R}^{N \times N}$, with ones on the sub and super diagonals and zero elsewhere. The authors then modify this adjacency matrix with a matrix $B \in \mathbb{R}^{M \times M}$ such that the $i, j$-th entry of $B$ modifies the $r_i, r_j$-th entry of $A$. This set up allows the authors to consider the case of $M$ modifications to the underlying string of nodes.

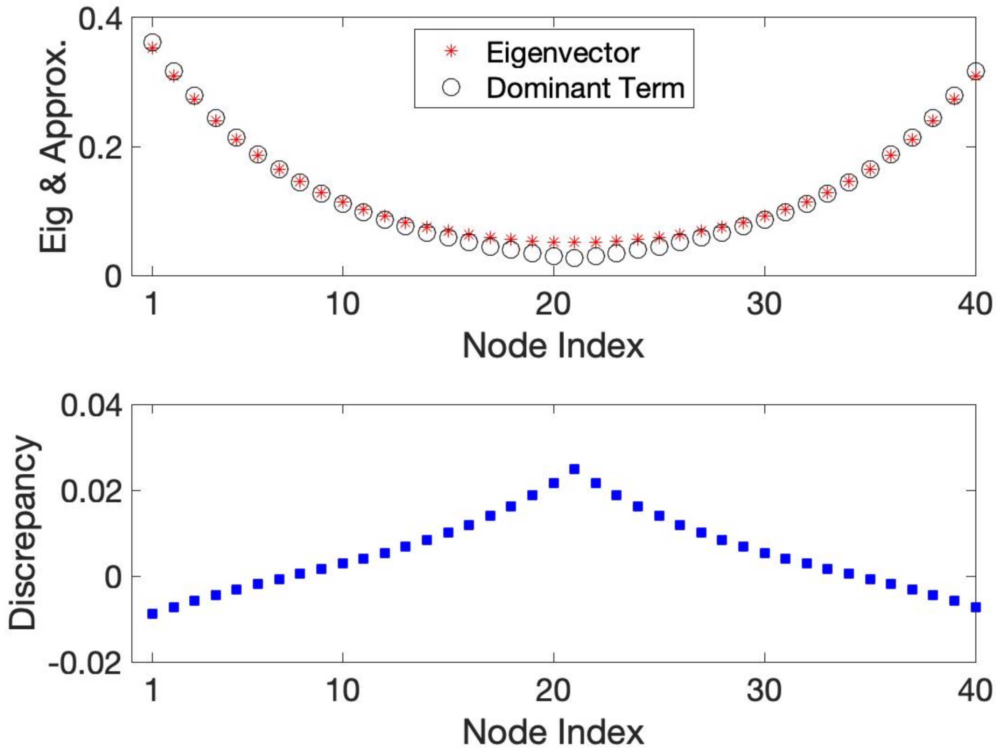



Figure 4.3: Upper picture: asterisks show components of the Perron-Frobenius vector and circles show the approximation $s_1^{p(i)}$ from (4.22), with $s_1$ defined in (4.23). Both vectors are normalized to have unit Euclidean norm. Lower picture: the discrepancy. From Theorem 4.2, $x_i - s_1^{p(i)}$ has the form $s_1^{N-p(i)}$, and hence grows geometrically away from the shortcut node.

In the simplest form, the authors consider the case of a modification of $A$ in position $(r, r)$ which represents a change in this position from 0 to $b$. The 'new' eigenvalue ($\lambda$) is shown to have increased and moved outwith the range $[-2, 2]$ (the spectrum of the unmodified matrix) which still contains all other eigenvalues. The components of the Perron-Frobenius eigenvector are shown to be almost zero except near node $r$ where there is a localised spike with geometric decay. Fitting an ansaztz of $x_j = t^{|j-r|}$ the authors approximate the value of $t$ and $\lambda$ by considering the system at the general nodes and node $r$ whilst ignoring the boundary effects which are argued to be of $O(t^L)$ where $L = \min(r-1, N-r)$. This gives the approximations

$$\begin{aligned} t &= \frac{1}{2}\left[-b + \text{sign}(b)\sqrt{4+b^2}\right] \\ \lambda &= \text{sign}(b)\sqrt{4+b^2}. \end{aligned}$$

The authors extend this approach to the case of $M$ modifications highlighting that it would be reasonable to observe $M$ localised spikes in the Perron-Frobenius eigenvector. The ansatz is now a sum of $M$ spikes centred at $r_k$ each with differing height $h_k$, i.e., $x_j = \sum_{k=1}^{M} h_k t^{|j-r_k|}$. Once more the authors ignore boundary effects and the higher order terms observed at row $r_k$ to give the approximations

$$\begin{aligned} t &= \frac{1}{2}\left[-\mu + \text{sign}(\mu)\sqrt{4+\mu^2}\right] \\ \lambda &= \text{sign}(\mu)\sqrt{4+\mu^2}, \end{aligned}$$

which relate the rate of decay $t$ and new eigenvalue $\lambda$ to $\mu$ (the leading eigenvalue of the modification matrix $B$). Similarly, the heights of the spikes are shown to be given by eigenvectors of the modification matrix $B$. The authors then prove that the approximations for the eigenvalue and eigenvector are within $O(t^L)$ to the actual eigenvalue-eigenvector pair.

We extend this approach in order to obtain an explicit expression for the Perron-Frobenius eigenvector. We note here that our problem is posed on a periodic ring unlike that of Liu, Strang and Ott who considered a string. Rather than ignore higher order contributions at the boundary or on spike rows we account for them explicitly and this allows us to identify geometrically increasing but exponentially small (in $N$) growth that was absorbed into the $O(t^L)$ term in [37]. Hence, Theorem 4.2 gives both exact solutions and approximations that are highly accurate in a large network limit. The technique of proof is similar to that in Theorem 4.1. However, we point out that whereas the Katz parameter $\alpha$ in Theorem 4.1 is fixed, the Perron-Frobenius eigenvalue $\lambda$ in Theorem 4.2 is dependent upon $N$.

**Theorem 4.2.** *For the undirected ring plus directed shortcut network with adjacency matrix $A = C + E$, let $\lambda$ denote the Perron-Frobenius eigenvalue. Then, for sufficiently large $N$, we have $2 < \lambda < 5/2$ and the Perron-Frobenius eigenvector $x$ has the form*

$$x_i = s_1^{p(i)} + s_2^{p(i)-N}, \tag{4.22}$$

*where $s_1, s_2$ are the roots of the palindromic quadratic $s^2 - \lambda s + 1$, so that*

$$s_1 = \frac{\lambda - \sqrt{\lambda^2 - 4}}{2}, \quad s_2 = \frac{\lambda + \sqrt{\lambda^2 - 4}}{2}, \tag{4.23}$$

*with $s_2 = 1/s_1$ and $0 < s_1 < 1 < s_2$.*

*Proof.* Following the proof of Theorem 4.1, we start with the ansatz

$$x_i = b + hs^{p(i)} \tag{4.24}$$

for the eigenvector, and attempt to satisfy the requisite equations for general node $k$, for node 1 and for the node, $q$, furthest away from node 1.

At a general node $k$, that is, in the $k$th row of (3.4), for $k \neq 1$ and $k \neq q$ we have

$$x_{k-1} - \lambda x_k + x_{k+1} = 0. \tag{4.25}$$

Inserting the ansatz (4.24) this becomes

$$b(2 - \lambda) + hs^{p(k)}(1/s + s - \lambda) = 0.$$

This may be satisfied by setting $b$ to zero and $s$ to the value $s_1$ or $s_2$. By linearity, we therefore continue with

$$x_i = h_1 s_1^{p(i)} + h_2 s_2^{p(i)}. \tag{4.26}$$

Since the eigenvector is only unique up to a scaling, we set $h_1 = 1$, leaving $h_2$ and $\lambda$ as free parameters. The two remaining equations to satisfy arise from nodes 1 and $q$ where

$$-\lambda x_1 + x_2 + x_L + x_N = 0, \tag{4.27}$$

and

$$\begin{cases} x_{N/2} - \lambda x_{N/2+1} + x_{N/2+2} = 0, \text{ if } N \text{ is even} & (4.28a) \\ x_{(N-1)/2} - \lambda x_{(N+1)/2} + x_{(N+3)/2} = 0, \text{ if } N \text{ is odd} & (4.28b) \end{cases}$$

respectively. In (4.28b) above, we have used $q = (N+1)/2$ but $q = (N+3)/2$ is equally valid and will result in the same analysis. Using (4.26) with $h_1 = 1$ we have

$$(2s_1 - \lambda + s_1^{L-1}) + h_2(2s_2 - \lambda + s_2^{L-1}) = 0, \tag{4.29}$$

and

$$\begin{cases} s_1^{N/2}(2s_2 - \lambda) + h_2 s_2^{N/2}(2s_1 - \lambda) = 0, \text{ if } N \text{ is even} & (4.30a) \\ s_1^{(N-1)/2}(s_2 + 1 - \lambda) + h_2 s_2^{(N-1)/2}(s_1 + 1 - \lambda) = 0, \text{ if } N \text{ is odd} & (4.30b) \end{cases}$$

respectively. Since $s_1 = 1/s_2$ and $s_1 + s_2 = \lambda$, multiplying (4.30a) by a factor of $s_2^{N/2}$ we obtain

$$h_2 = s_2^{-N}\left(\frac{-(2s_2 - \lambda)}{2s_1 - \lambda}\right) = s_2^{-N}\left(\frac{-(s_2 - s_1)}{s_1 - s_2}\right) = s_1^N.$$

Similarly, multiplying (4.30b) by a factor of $s_2^{(N-1)/2}$ we obtain

$$h_2 = s_2^{-(N-1)}\left(\frac{-(s_2 + 1 - \lambda)}{s_1 + 1 - \lambda}\right) = s_2^{-(N-1)}\left(\frac{-s_1(s_2 - 1)}{1 - s_2}\right) = s_1^N.$$

We have shown that both (4.30a) and (4.30b) reduce to $h_2 = s_1^N$ so writing (4.29) in terms of the remaining unknown, $\lambda$, we have

$$F(\lambda) = 0, \tag{4.31}$$

where

$$\begin{aligned} F(\lambda) &= (s_1 - s_2)(1 - s_1^N) + s_1^{L-1} + s_1^{N-L+1} \\ &= -\sqrt{\lambda^2 - 4}\left(1 - \left(\frac{\lambda - \sqrt{\lambda^2-4}}{2}\right)^N\right) + \left(\frac{\lambda - \sqrt{\lambda^2-4}}{2}\right)^{L-1} \\ &\quad + \left(\frac{\lambda - \sqrt{\lambda^2-4}}{2}\right)^{N-L+1}. \end{aligned} \tag{4.32}$$

It is straightforward to show that $F(2) > 0$, whereas at $\lambda = 5/2$ we have

$$F(5/2) = -\frac{3}{2}\left(1 - 2^{-N}\right) + 2^{1-L} + 2^{L-N-1}$$

for large $N$.

In the $N$ even case we have $L = \theta(N/2 + 1)$ so $1 - L = 1 - \theta - \theta N/2$ and $L - N - 1 = \theta N/2 + \theta - N - 1 = -(2-\theta)N/2 - (1-\theta)$. Therefore,

$$\begin{aligned} F(5/2) &= -\frac{3}{2}\left(1 - 2^{-N}\right) + 2^{1-L} + 2^{L-N-1} \\ &= -\frac{3}{2} + 2^{1-L} + 2^{L-N-1} + \frac{3}{2}2^{-N} \\ &= -\frac{3}{2} + 2^{1-\theta}2^{-\theta N/2} + 2^{-(1-\theta)}2^{-(2-\theta)N/2} + \frac{3}{2}2^{-N} \\ &= -\frac{3}{2} + O(2^{-\theta N/2}) + O(2^{-(2-\theta)N/2}) < 0, \end{aligned}$$

for large $N$.

In the $N$ odd case we have $L = \theta(N+1)/2$ so $1 - L = 1 - \theta/2 - \theta N/2$ and

$L - N - 1 = \theta N/2 + \theta/2 - N - 1 = -(2-\theta)N/2 - (1-\theta/2)$. Therefore,

$$
\begin{aligned}
F(5/2) &= -\frac{3}{2}\left(1 - 2^{-N}\right) + 2^{1-L} + 2^{L-N-1} \\
&= -\frac{3}{2} + 2^{1-L} + 2^{L-N-1} + \frac{3}{2}2^{-N} \\
&= -\frac{3}{2} + 2^{1-\theta/2}2^{-\theta N/2} + 2^{-(1-\theta/2)}2^{-(2-\theta)N/2} + \frac{3}{2}2^{-N} \\
&= -\frac{3}{2} + O(2^{-\theta N/2}) + O(2^{-(2-\theta)N/2}) < 0,
\end{aligned}
$$

for large $N$.

So the continuous function $F$ changes sign in the interval $(2, 5/2)$.

We have thus established that the nonsymmetric matrix $C+E$ has an eigenvector of the form (4.22) with an eigenvalue in $(2, 5/2)$. We can rule out the possibility of an eigenvector existing that has a larger eigenvalue. This follows from [30, Problem 8.4.P15] (which applies to all nonnegative, irreducible matrices)—because we have constructed an eigenvector with all $x_i > 0$, it must be a Perron-Frobenius eigenvector. □

## 4.5 Bounds on the Perron-Frobenius eigenvalue

In Theorem 4.2 it was sufficient to have a crude bound on the associated Perron-Frobenius eigenvalue. Here, we derive a sharper result that justifies the restriction $0 < \alpha < 1/2$ imposed in Section 4.3.

**Theorem 4.3.** *Given any $\epsilon > 0$ and $K > 0$, for sufficiently large $N$ the Perron-Frobenius eigenvalue $\lambda$ in Theorem 4.2 is such that*

$$
2 + \frac{K}{N^2} < \lambda < 2 + \frac{K}{N^{2-\epsilon}}.
$$

*Proof.* We consider $F(\lambda)$ in (4.32) for values of $\lambda$ close to 2, noting that the Perron-Frobenius eigenvalue satisfies $F(\lambda) = 0$. That is we regard $\lambda$ as a general point and use the idea that if $F(\lambda_a)F(\lambda_b) < 0$ then $F(\lambda) = 0$ for some $\lambda \in (\lambda_a, \lambda_b)$.

If $\lambda$ has the form

$$\lambda = 2 + KN^{-\beta}$$

for some $\beta > 0$, then to leading order in (4.23) we have

$$s_1 = 1 - K^{1/2}N^{-\beta/2} + O(N^{-\beta}), \tag{4.33}$$

$$s_2 = 1 + K^{1/2}N^{-\beta/2} + O(N^{-\beta}). \tag{4.34}$$

**Now suppose** $\beta = 2$, then recalling that $L$ is proportional to $N$, we have

$$\begin{aligned}
\log s_1^{L-1} &= (L-1)\log\left(1 - K^{1/2}N^{-1} + O(N^{-2})\right) \\
&= (L-1)\left(-K^{1/2}N^{-1}\right) + O(N^{-2}) \\
&\to -\gamma K^{1/2}, \\
\log s_1^{N-L+1} &= (N-L+1)\log\left(1 - K^{1/2}N^{-1} + O(N^{-2})\right) \\
&= (N-L+1)\left(-K^{1/2}N^{-1}\right) + O(N^{-2}) \\
&\to -(1-\gamma)K^{1/2}, \\
\log s_1^{N} &= N\log\left(1 - K^{1/2}N^{-1} + O(N^{-2})\right) \\
&= N\left(-K^{1/2}N^{-1}\right) + O(N^{-2}) \\
&\to -K^{1/2},
\end{aligned}$$

for some fixed $\gamma = (L-1)/N > 0$. So

$$s_1^{L-1} \to e^{-\gamma K^{1/2}}, \qquad s_1^{N-L+1} \to e^{-(1-\gamma)K^{1/2}} \quad s_1^{N} \to e^{-K^{1/2}}.$$

Since $F(\lambda)$ in (4.32) may be written

$$F(\lambda) = (2s_1 - \lambda)(1 - s_1^N) + s_1^{L-1} + s_1^{N-L+1},$$

we conclude that it takes the form of an asymptotically small negative term plus positive terms that are bounded away from zero. Hence $F(\lambda) > 0$ for large $N$.

**Now suppose** $\beta < 2$**.** Similar analysis shows that, as $N \to \infty$,

$$s_1^{L-1} \to 0, \qquad s_1^{N-L+1} \to 0, \qquad s_1^{N} \to 0,$$

at rates that are exponential; that is, as fast as $e^{-\nu N^{1-\beta/2}}$ for some fixed $\nu > 0$. It follows that $F(\lambda)$ in (4.32) is dominated by the term $-2\sqrt{K}N^{-\beta/2}$, so $F(\lambda) < 0$ for sufficiently large $N$.

$\square$

# Chapter 5

# Extensions

In this Chapter we consider variations of the periodic ring plus shortcut across the network by examining the case of (a) multiple shortcuts and (b) nodes connecting to more than one neighbour on each side. To do so, we will first consider the simplest form of such a network and then conjecture to the general case.

In this Chapter we only consider Katz centrality as this work forms a basis for our approach to more complex examples presented later in Chapter 6.

## 5.1 $M$-shortcuts across the ring

In this section we consider the case of adding more than one shortcut across the network. To keep things simple and identify the key ideas we will consider the case of adding one undirected shortcut which we can consider to be a pair of directed shortcuts. We then hypothesise about what we would expect to happen by removing the constraint of the additional shortcut pointing backwards at itself and then conjecture about $M$ shortcuts across the network. We also test our conjecture with computational experiments.

Following on from Chapter 4 we assume that node 1 and node $L$ are involved in the shortcuts. We therefore have that node 1 has a shortcut to node $L$ and node $L$ has a shortcut to node 1. Using the same notation, we have the adjacency matrix $A = C + E'$ where $E'(1, L) = E'(L, 1) = 1$. Without loss of generality, we assume that $L$ is not beyond the half point. Therefore for some fixed proportion $0 < \theta \leq 1$

$$L = \begin{cases} \lfloor \theta(N/2 + 1) \rfloor & \text{when } N \text{ is even,} \\ \lfloor \theta(N + 1)/2 \rfloor & \text{when } N \text{ is odd.} \end{cases} \tag{5.1}$$

As was the case in Chapter 4 we let $p(i)$ denote the periodic distance from node $i$ to node 1 as in (4.3) and let $p_L(i)$ denote the periodic distance from node $i$ to node $L$, i.e.

$$p_L(i) = \min\left(|i - L|, |N - (i - L)|\right). \tag{5.2}$$

In the upper picture of Figure 5.1, the asterisks show Katz centrality values for a network with adjacency matrix $A = C + E'$, that is, components of $x$ from (3.1). To make the key effects visible we chose a small network size. We used an $N = 20$ node ring with a shortcut from node 1 to node $L = 8$ and a shortcut from node $L = 8$ to node 1, for $\alpha = 0.3$. Since node 1 and node $L$ own an extra, long-range edge, they attain the highest centrality scores and hence are equally ranked. The more distant nodes, from both node 1 and node $L$, which in this case would be node 14 and node 15 should be deemed least central. As we have shown previously, in the case of a single shortcut, the components of $x_i$ when suitably shifted were varying geometrically. In this instance, we have 2 shortcuts and therefore we might expect when the components of $x_i$, when suitably shifted, to vary as a sum of 2 geometric terms as we move around the ring, that is $x_i = b + h(t^{p(i)} + t^{p_L(i)})$. Using an ansatz of this form leads to the circles in the upper picture of Figure 5.1. The agreement between the two are close — below $5 \times 10^{-5}$ in Euclidean norm.

The lower picture of Figure 5.1 shows, on a log-scale, the discrepancy between those asterisks (true solution) and circles (geometrically decaying ansatz). We observe that this discrepancy is at its maximum at nodes furthest periodically from those nodes that own an additional shortcut, i.e. nodes 11 and 18. The discrepancy is a very small contribution that *decays* geometrically as we move away periodically from nodes 11 and 18.

We now state a theorem that explains Figure 5.1. This theorem concerns the limit $N \to \infty$ with a fixed Katz parameter $0 < \alpha < 1/\rho(A)$.

**Theorem 5.1.** *For the undirected ring plus undirected shortcut network with adjacency matrix $A = C + E'$, for sufficiently large $N$ the unique solution of the Katz system (3.1) has the form*

$$x_i = b + h_1 \left( t_1^{p(i)} + t_1^{p_L(i)} \right) + h_2 \left( t_2^{p(i)} + t_2^{p_L(i)} \right). \tag{5.3}$$

*Here, $b, t_1, t_2, h_1, h_2$ are constants, i.e., independent of $i$, and $b, t_1, t_2$ are also independent of $N$. In particular, $h_1 > 0$, $h_2 > 0$, $b = 1/(1 - 2\alpha)$ and $t_1, t_2$ are the roots of the palindromic quadratic $\alpha t^2 - t + \alpha$, so that*

$$t_1 = \frac{1 - \sqrt{1 - 4\alpha^2}}{2\alpha}, \quad t_2 = \frac{1 + \sqrt{1 - 4\alpha^2}}{2\alpha}, \tag{5.4}$$

*with $t_2 = 1/t_1$ and $0 < t_1 < 1 < t_2$. Moreover, the terms in (5.3) involving $h_2$ are exponentially small asymptotically, in the sense that,*

$$x_i = b + h_1 \left( t_1^{p(i)} + t_1^{p_L(i)} \right) + O(t_1^{N/2}), \tag{5.5}$$

*with $h_1 = O(1)$.*

*Proof.* Our proof proceeds in a similar way to the proof of Theorem 4.1. We begin by assuming that the height of the spikes at nodes 1 and $L$ could be different and use the ansatz

$$x_i = b + h t^{p(i)} + h' t^{p_L(i)}. \tag{5.6}$$

The Katz system (3.1) essentially reduces to five scalar equations where we need to consider:

- General node $k$ where each neighbouring node differs by a factor of $t$.
- Node 1 which has a shortcut across the ring to node $L$.
- Node $L$ which has a shortcut across the ring to node 1.
- The node(s) furthest away from node 1 periodically, i.e.

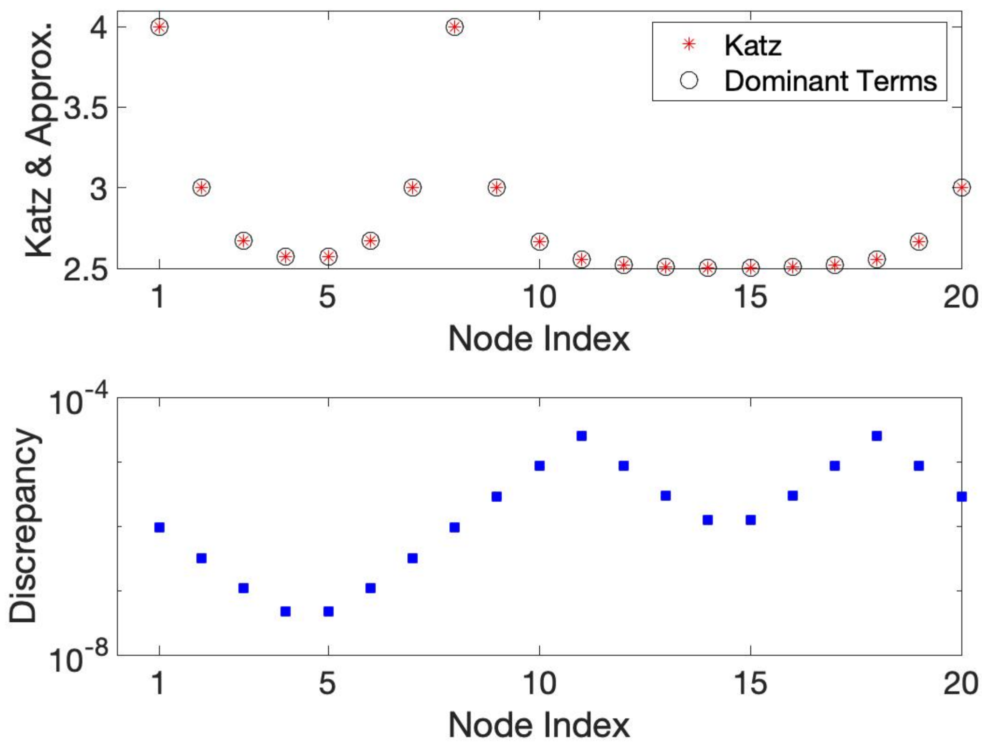



Figure 5.1: Upper picture: asterisks show components of Katz vector from (3.1) and circles show the approximation $b + h_1(t_1^{p(i)} + t_1^{p_L(i)})$ from (5.5). Here $b = 1/(1-2\alpha)$, $t_1$ is defined in (5.4) and $h_1$ was found by solving (5.18). Lower picture: the discrepancy $x_i - b - h_1(t_1^{p(i)} + t_1^{p_L(i)})$ on a log scale. From Theorem 5.1, this quantity has the form $h_2(t_2^{p(i)} + t_2^{p_L(i)})$, and hence grows geometrically away from the shortcut nodes. However, it is uniformly $O(t_1^{N/2})$ for a fixed $0 < t_1 < 1$, and hence rapidly becomes negligible as the network size $N$ increases.

- node $q_1 = N/2 + 1$, if $N$ is even,
- node $q_1 = (N+1)/2$ or $q_1 = (N+3)/2$, if $N$ is odd.

- The node(s) furthest away from node $L$ periodically, i.e.
  - node $q_2 = N/2 + L$, if $N$ is even,
  - node $q_2 = (N-1)/2 + L$ or $q_2 = (N+1)/2 + L$, if $N$ is odd.

At a general node $k$, corresponding to the $k$th row of the linear system, for $k \neq 1$, $k \neq L$, $k \neq q_1$ and $k \neq q_2$ we have

$$-\alpha x_{k-1} + x_k - \alpha x_{k+1} = 1. \tag{5.7}$$

Inserting (5.6), this becomes

$$b(1-2\alpha) + ht^{p(k)}\left[1 - \alpha t - \alpha/t\right] + h' t^{p_L(k)}\left[1 - \alpha t - \alpha/t\right] = 1.$$

We can satisfy this equation independently of $k$ by setting $b = 1/(1-2\alpha)$ and choosing $t$ to be either root of the quadratic $\alpha t^2 - t + \alpha$. By linearity of (5.7), we may extend (5.6) to include a linear combination involving both roots, so our ansatz becomes

$$x_k = b + h_1 t_1^{p(k)} + h_2 t_2^{p(k)} + h_1' t_1^{p_L(k)} + h_2' t_2^{p_L(k)}, \tag{5.8}$$

where $t_1$ and $t_2$ are given in (5.4).

In addition to the useful facts (4.11) and (4.12) in Section 4.3 we also note that

$$1 - 2\alpha t_1 - \alpha = \sqrt{1 - 4\alpha^2} - \alpha > 0 \tag{5.9}$$

on our domain $0 < \alpha < 1/\rho(A)$.

The four remaining equations to satisfy from (3.1) arise at nodes 1, $L$, $q_1$ and $q_2$, where we require

$$-\alpha x_N + x_1 - \alpha x_2 - \alpha x_L \ = \ 1, \tag{5.10}$$

$$-\alpha x_{L-1} + x_L - \alpha x_{L+1} - \alpha x_1 \ = \ 1, \tag{5.11}$$

$$\begin{cases} -\alpha x_{N/2} + x_{N/2+1} - \alpha x_{N/2+2} = 1, & \text{if } N \text{ is even} \quad (5.12a) \\ -\alpha x_{(N-1)/2} + x_{(N+1)/2} - \alpha x_{(N+3)/2} = 1, & \text{if } N \text{ is odd} \quad (5.12b) \end{cases}$$

and

$$\begin{cases} -\alpha x_{N/2+L-1} + x_{N/2+L} - \alpha x_{N/2+L+1} = 1, & \text{if } N \text{ is even} \quad (5.13a) \\ -\alpha x_{(N-3)/2+L} + x_{(N-1)/2+L} - \alpha x_{(N+1)/2+L} = 1, & \text{if } N \text{ is odd} \quad (5.13b) \end{cases}$$

respectively. In (5.12b) above, we have used $q_1 = (N+1)/2$ but $q_1 = (N+3)/2$ is equally valid and will result in the same conclusion. Similarly, in (5.13b) above, we have used $q_2 = (N-1)/2 + L$ but $q_2 = (N+1)/2 + L$ is equally valid and will result in the same conclusion.

Inserting (5.8) into (5.10) – (5.13b), using $b(1-2\alpha) = 1$ and $1 - \alpha t - \alpha/t = 0$, we have

$$h_1 \left[1 - 2\alpha t_1 - \alpha t_1^{L-1}\right] + h_2 \left[1 - 2\alpha t_2 - \alpha t_2^{L-1}\right] - \alpha h_1' - \alpha h_2' = b\alpha, \quad (5.14)$$

$$h_1' \left[1 - 2\alpha t_1 - \alpha t_1^{L-1}\right] + h_2' \left[1 - 2\alpha t_2 - \alpha t_2^{L-1}\right] - \alpha h_1 - \alpha h_2 = b\alpha, \quad (5.15)$$

$$\begin{cases} h_1 t_1^{N/2-1} \left[t_1 - 2\alpha\right] + h_2 t_2^{N/2-1} \left[t_2 - 2\alpha\right] = 0, & \text{if } N \text{ is even} \quad (5.16a) \\ h_1 t_1^{(N-3)/2} \left[(1-\alpha)t_1 - \alpha\right] + h_2 t_2^{(N-3)/2} \left[(1-\alpha)t_2 - \alpha\right] = 0, & \text{if } N \text{ is odd} \quad (5.16b) \end{cases}$$

and

$$\begin{cases} h_1' t_1^{N/2-1} \left[t_1 - 2\alpha\right] + h_2' t_2^{N/2-1} \left[t_2 - 2\alpha\right] = 0, & \text{if } N \text{ is even} \quad (5.17a) \\ h_1' t_1^{(N-3)/2} \left[(1-\alpha)t_1 - \alpha\right] + h_2' t_2^{(N-3)/2} \left[(1-\alpha)t_2 - \alpha\right] = 0, & \text{if } N \text{ is odd} \quad (5.17b) \end{cases}$$

We can now conclude that $h = h'$, that is $h_1 = h_1'$ and $h_2 = h_2'$, meaning that we can now reduce the system to only considering the situation at nodes 1 and $q_1$ or nodes $L$ and $q_2$.

**Consider the case where the number of nodes, $N$, is even.**

The linear system is

$$\begin{bmatrix} 1 - \alpha - 2\alpha t_1 - \alpha t_1^{L-1} & 1 - \alpha - 2\alpha t_2 - \alpha t_2^{L-1} \\ t_1^{N/2-1}(t_1 - 2\alpha) & t_2^{N/2-1}(t_2 - 2\alpha) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} b\alpha \\ 0 \end{bmatrix}. \quad (5.18)$$

To prove that a solution of the form (5.8) exists, we now show that this system is nonsingular. The determinant may be written

$$\left(1-\alpha-2\alpha t_1-\alpha t_1^{L-1}\right)t_2^{N/2}(1-2\alpha t_1)-\left(1-\alpha-2\alpha t_2-\alpha t_2^{L-1}\right)t_1^{N/2}(1-2\alpha t_2). \tag{5.19}$$

Recalling that $0<t_1<1<t_2$, we see from (4.11) and (5.9) that, as $N\to\infty$ and hence $L\to\infty$, the first term in (5.19) becomes, to leading order,

$$(1-2\alpha t_1-\alpha)\,t_2^{N/2}\,(1-2\alpha t_1)\to\infty.$$

The second term in (5.19) becomes, to leading order,

$$\alpha t_2^{L-1}t_1^{N/2-1}(t_1-2\alpha)=\alpha t_1^{N/2-L}(t_1-2\alpha). \tag{5.20}$$

Recall that $L=\lfloor\theta(N/2+1)\rfloor$, so

$$N/2-L>N/2-\theta(N/2+1)=(1-\theta)N/2-\theta.$$

Hence, as $N\to\infty$, term $\alpha t_1^{N/2-L}(t_1-2\alpha)$ in (5.20) is bounded for $\theta=1$, and tends to zero for $\theta<1$. So the first term in (5.19) dominates, and the determinant is bounded away from zero for large $N$.

**Consider the case where the number of nodes, $N$, is odd.**

We have the linear system

$$\begin{bmatrix} 1-\alpha-2\alpha t_1-\alpha t_1^{L-1} & 1-\alpha-2\alpha t_2-\alpha t_2^{L-1} \\ t_1^{(N-3)/2}\left[(1-\alpha)t_1-\alpha\right] & t_2^{(N-3)/2}\left[(1-\alpha)t_2-\alpha\right] \end{bmatrix}\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}=\begin{bmatrix} b\alpha \\ 0 \end{bmatrix}. \tag{5.21}$$

To prove that a solution of the form (5.8) exists, we again show that this system is nonsingular. The determinant may be written

$$\begin{aligned}&\left(1-\alpha-2\alpha t_1-\alpha t_1^{L-1}\right)t_2^{(N-3)/2}\left[(1-\alpha)t_2-\alpha\right]\\ -&\left(1-\alpha-2\alpha t_2-\alpha t_2^{L-1}\right)t_1^{(N-3)/2}\left[(1-\alpha)t_1-\alpha\right].\end{aligned} \tag{5.22}$$

Recalling that $0<t_1<1<t_2$, we see from (4.12) and (5.9) that, as $N\to\infty$ and

hence $L \to \infty$, the first term in (5.22) becomes, to leading order,

$$(1 - 2\alpha t_1 - \alpha)\, t_2^{(N-3)/2} \left[(1-\alpha)t_2 - \alpha\right] \to \infty.$$

The second term in (5.22) becomes, to leading order,

$$\alpha t_2^{L-1} t_1^{(N-3)/2} \left[(1-\alpha)t_1 - \alpha\right] = \alpha t_1^{(N-1)/2-L} \left[(1-\alpha)t_1 - \alpha\right]. \tag{5.23}$$

Recall that $L = \lfloor \theta((N+1)/2) \rfloor$, so

$$(N-1)/2 - L > (N-1)/2 - \theta((N+1)/2) = (1-\theta)(N-1)/2 - \theta.$$

Hence, as $N \to \infty$, term $\alpha t_1^{(N-1)/2-L} \left[(1-\alpha)t_1 - \alpha\right]$ in (5.23) is bounded for $\theta = 1$, and tends to zero for $\theta < 1$. So the first term in (5.22) dominates, and the determinant is bounded away from zero for large $N$.

**Continuation of general case: $N$ even or odd.**

We have shown that for sufficiently large $N$ the unique solution of the Katz system (3.1) has the form (5.8). We now follow up by showing that, in the exact solution (5.8), the growing terms $h_2 t_2^{p(i)}$ and $h_2 t_2^{p_L(i)}$ are negligible for large $N$.

To do so, we first show that the second equations in (5.18) and (5.21) can both be reduced to the same condition. Using (5.18) and (4.11),

$$h_2 = -\frac{t_1^{N/2-1}(t_1 - 2\alpha)}{t_2^{N/2-1}(t_2 - 2\alpha)} h_1 = -t_1^N \frac{1 - 2\alpha t_2}{1 - 2\alpha t_1} h_1 = h_1 t_1^N. \tag{5.24}$$

Similarly, using (5.21) and (4.12),

$$h_2 = -\frac{t_1^{(N-3)/2}\left[(1-\alpha)t_1 - \alpha\right]}{t_2^{(N-3)/2}\left[(1-\alpha)t_2 - \alpha\right]} h_1 = -t_1^{N-3} \frac{t_1^2(1-t_2)}{t_2(t_2-1)} h_1 = h_1 t_1^N. \tag{5.24}$$

Hence, in the first equation of (5.18) and (5.21),

$$h_1 \left[1 - \alpha - 2\alpha t_1 - \alpha t_1^{L-1} - \alpha t_1^{N-L+1} - 2\alpha t_1^{N-1} + (1-\alpha)t_1^N\right] = b\alpha.$$

We see that $h_1 = O(1)$. Also, using (5.9), we deduce that $h_1 > 0$ for large

$N$. It then follows from (5.24) that $h_2 > 0$ and $h_2 = O(t_1^N)$. Since the term $h_2 t_2^{p(k)} = h_1 t_1^{N-p(k)}$ is largest when $k = N/2+1$ and $h_2 t_2^{p_L(k)} = h_1 t_1^{N-p_L(k)}$ is largest when $k = N/2+L$ , we also see that

$$\begin{aligned} x_i &= b + h_1 \left( t_1^{p(i)} + t_1^{p_L(i)} \right) + h_1 \left( t_1^{N-p(i)} + t_1^{N-p_L(i)} \right) \\ &= b + h_1 \left( t_1^{p(i)} + t_1^{p_L(i)} \right) + O(t_1^{N/2}). \end{aligned}$$

□

For a large, fixed, value of $N$, the expression (5.5) in Theorem 5.1 will be dominated by the contribution from whichever of nodes 1 and $L$ is closest to node $i$. Hence, the Katz centrality of node $i$ will take the approximate form

$$\begin{aligned} x_i &= b + h_1 t_1^{p_{1,L}(i)} + O(t_1^{L-1}) + O(t_1^{N/2}), \\ &= b + h_1 t_1^{p_{1,L}(i)} + O(t_1^{\theta(N/2+1)-1}) + O(t_1^{N/2}), \\ &= b + h_1 t_1^{p_{1,L}(i)} + O(t_1^{\theta N/2}) + O(t_1^{N/2}), \end{aligned}$$

where

$$p_{1,L}(i) = \min\left(|i-1|, |N-(i-1)|, |i-L|, |N-(i-L)|\right).$$

We began our proof by assuming that the height of the spikes at the originating shortcut nodes could be different and were then able to show that $h = h'$, i.e., that the height of the spikes were equal. This was only the case in our proof as we considered an undirected shortcut (a pair of directed shortcuts that point in opposite directions) and therefore they span the same periodic distance across the ring.

Therefore, considering the condition of a pair of independent shortcuts (shortcuts that do not involve the same pair of nodes) then retaining the definition of a shortcut from node 1 to node $L$ we would impose a similar definition to the second shortcut. That is, we have an additional shortcut from node $S$ to node $T$ whereby node $S$ is no more than half way around the network, i.e., $S < N/2+1$ and node $T$ does not extend more than half-way around the ring from node $S$.

Therefore for some fixed proportion $0 < \phi \leq 1$

$$T = \begin{cases} \lfloor \phi(N/2+1) \rfloor + S - 1 & \text{when } N \text{ is even,} \\ \lfloor \phi(N+1)/2 \rfloor + S - 1 & \text{when } N \text{ is odd.} \end{cases} \tag{5.25}$$

Looking back at (5.14) – (5.15) we would expect the height of the spikes to be slightly different unless the shortcuts still span the same periodic distance. However, in the asymptotic regime when $N \to \infty$ and hence $L \to \infty$ and $T \to \infty$ the height of the spikes should be equal regardless.

We therefore conjecture that in the general case of a nearest neighbour periodic ring with $M$ well-separated shortcuts across the network originating from nodes $y_1, y_2, \ldots, y_m$ the solution of the Katz system (3.1) is given by

$$x_i = b + \sum_{j=1}^{M} h_j t_1^{p_{y_j}(i)} + t_1^N \sum_{j=1}^{M} h_j t_2^{p_{y_j}(i)}.$$

Here, $b, t_1, t_2, h_j$ are constants, i.e., independent of $i$, and $b, t_1, t_2$ are also independent of $N$. In particular, $h_j > 0$, $b = 1/(1-2\alpha)$ and $t_1, t_2$ are the roots of the palindromic quadratic $\alpha t^2 - t + \alpha$ so that $t_1$ and $t_2$ are as provided in (5.4) with $t_2 = 1/t_1$ and $0 < t_1 < 1 < t_2$. Moreover, as $N \to \infty$ the Katz system (3.1) has the approximate form

$$x_i \approx b + h \left( \sum_{j=1}^{M} t_1^{p_{y_j}(i)} \right). \tag{5.26}$$

We test this conjecture computationally by considering a nearest neighbour ring with 3 directed shortcuts. To allow comparison as we increase the number of nodes in the ring (keeping $N$ even) we define our shortcuts as $1 \mapsto L$, $S_1 \mapsto T_1$ and $S_2 \mapsto T_2$ where

$$L = \left\lfloor \frac{7}{10} \left( \frac{N}{2} + 1 \right) \right\rfloor,$$

$$S_1 = \left\lfloor \frac{1}{2} \left( \frac{N}{2} + 1 \right) \right\rfloor, \quad T_1 = \left\lfloor \frac{3}{5} \left( \frac{N}{2} + 1 \right) \right\rfloor + S_1 - 1,$$

$$S_2 = \left\lfloor \frac{9}{10} \left( \frac{N}{2} + 1 \right) \right\rfloor, \quad T_2 = \left\lfloor \frac{4}{5} \left( \frac{N}{2} + 1 \right) \right\rfloor + S_2 - 1.$$

Our shortcuts therefore scale and remain well-separated with the increase in the number of nodes. We also define the Katz parameter $\alpha = 0.4 < 1/\rho(A)$. Let $x$ be the solution by solving the Katz system (3.1) directly and let $\widehat{x}$ be our ansatz as defined in (5.26).

Figure 5.2 shows the Euclidean-norm of the residual error, $||r||_2 = ||x - \widehat{x}||_2$ on a log scale. We observe the residual error decreasing at what appears to be a constant rate until it reaches a plateau at around $N = 500$. The residual error when $N = 1,000$ is $< 2.5 \times 10^{-14}$ providing evidence to support the conjecture.

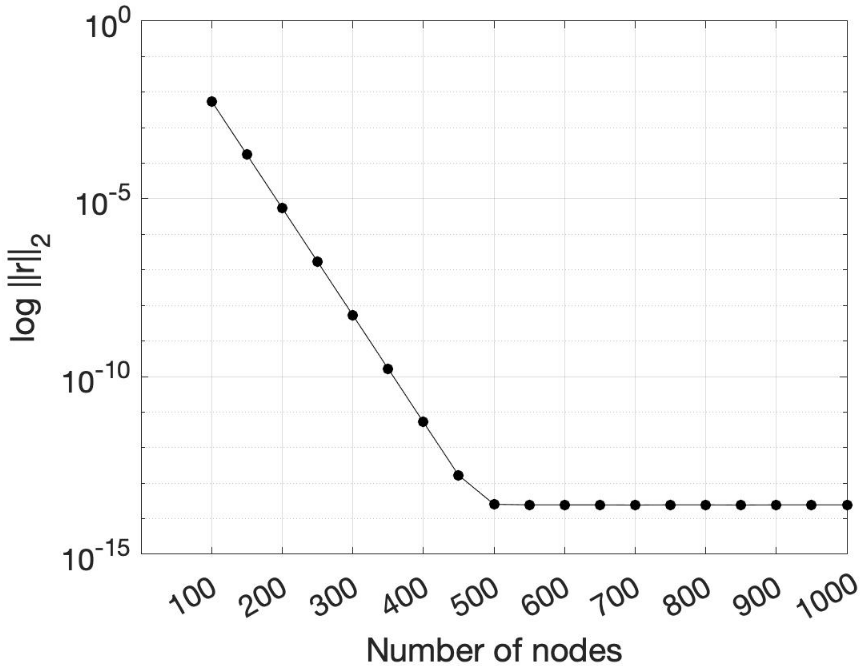



Figure 5.2: Euclidean-norm of the residual error $r = x - \widehat{x}$ on a log scale for the nearest neighbour periodic ring with 3 directed shortcuts from $1 \mapsto L$, $S_1 \mapsto T_1$ and $S_2 \mapsto T_2$

## 5.2 $2n$-nearest neighbours

In this section we consider the case of having a $2n$-nearest neighbour ring so that each node connects to the $n$ nearest neighbours on each side with a directed shortcut, as defined previously, from node 1 to node $L$. To keep things simple we first analyse the case of a 4 neighbour ring and then conjecture about what we would expect to happen in the general case of $2n$ neighbours.

Let $C_4 \in \mathbb{R}^{N\times N}$ be the adjacency matrix for the periodic 4-nearest neighbour on each side ring; so in the $N = 6$ case,

$$C_4 = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}. \tag{5.27}$$

As with $C$ in (4.2), we note that $C_4$ is symmetric and circulant, and standard theory shows that the Perron-Frobenius eigenvalue is $\rho(C_4) = 4$. We now add a single directed shortcut to the ring from node 1 to node $L$ as previously defined so that our adjacency matrix is $A = C_4 + E$.

In the upper picture of Figure 5.3, the asterisks show Katz centrality values for a network with adjacency matrix $A = C_4 + E$, that is, components of $x$ from (3.1). We chose a small network size in order to make the key effects visible. More precisely, we used an $N = 20$ node ring with a shortcut from node 1 to node $L = 8$, for $\alpha = 0.1$. Because node 1 owns the extra, long-range edge, it attains the highest centrality score, at around 1.84. The most distant node, periodically, node 11, is deemed the least central. An observable difference from that shown previously is that that nodes 3 and 19, which now connect to node 1 have had a boost in centrality. Insight from [37] and results from Chapter 4, suggests that components of $x_i$, when suitably shifted, might be varying as a sum of differing geometric components as the index $i$ moves periodically around the ring. We would expect to see one of the geometric components presenting as an oscillatory term in that it should increase the centrality for odd nodes and decrease the centrality for even nodes. That is, we have $x_i = b + h_1 t_1^{p(i)} + h_3 t_3^{p(i)}$ for some constants $b, h_1 > 0, h_3 > 0$, $0 < t_1 < 1$ and $-1 < t_3 < 0$. Fitting an ansatz of this form leads us to the circles in the upper picture of Figure 5.3. The agreement is close—below $9 \times 10^{-6}$ in Euclidean norm.

The lower picture in Figure 5.3 shows, on a log scale, the discrepancy between those asterisks (true solution) and circles (geometric decay ansatz). We see a very small contribution that, in contrast to the overall solution, *grows* geometrically as we move periodically away from node 1. We would once again expect to observe an oscillatory term that changes sign on even and odd nodes.

We now state a theorem that explains Figure 5.3. This theorem concerns the limit $N \to \infty$ with a fixed Katz parameter $0 < \alpha < 1/4$.

**Theorem 5.2.** *For the undirected* 4*-nearest neighbour ring plus undirected shortcut network with adjacency matrix* $A = C_4 + E$*, for sufficiently large* $N$ *the unique solution of the Katz system (3.1) has the form*

$$x_i = b + h_1 t_1^{p(i)} + h_2 t_2^{p(i)} + h_3 t_3^{p(i)} + h_4 t_4^{p(i)}. \tag{5.28}$$

*Here,* $b, t_1, t_2, t_3, t_4, h_1, h_2, h_3, h_4$ *are constants, i.e., independent of* $i$*, and* $b, t_1, t_2, t_3, t_4$ *are also independent of* $N$*. In particular,* $h_1 > 0$*,* $h_2 > 0$*,* $h_3 > 0$*,* $h_4 > 0$ *if* $N$ *is even and* $h_4 < 0$ *if* $N$ *is odd,* $b = 1/(1-2\alpha)$ *and* $t_1, t_2, t_3, t_4$ *are the roots of the palindromic quartic* $\alpha t^4 + \alpha t^3 - t^2 + \alpha t + \alpha$*, so that*

$$\begin{aligned} t_1 &= \frac{x - \sqrt{y - 2x} - 1}{4}, \quad t_2 = \frac{x + \sqrt{y - 2x} - 1}{4}, \\ t_3 &= \frac{-x + \sqrt{y + 2x} - 1}{4}, \quad t_4 = \frac{-x - \sqrt{y + 2x} - 1}{4}, \end{aligned} \tag{5.29}$$

*where* $x = \sqrt{4/\alpha + 9}$ *and* $y = 4/\alpha - 6$*. We have the roots appearing in reciprocal pairs such that* $t_2 = 1/t_1$*,* $t_4 = 1/t_3$ *and* $t_4 < -1 < t_3 < 0 < t_1 < 1 < t_2$*. Moreover, the terms in (5.28) involving* $h_2$ *and* $h_4$ *are exponentially small asymptotically, in the sense that,*

$$x_i = b + h_1 t_1^{p(i)} + h_3 t_3^{p(i)} + O(t_1^{N/2}) + O(|t_3|^{N/2}), \tag{5.30}$$

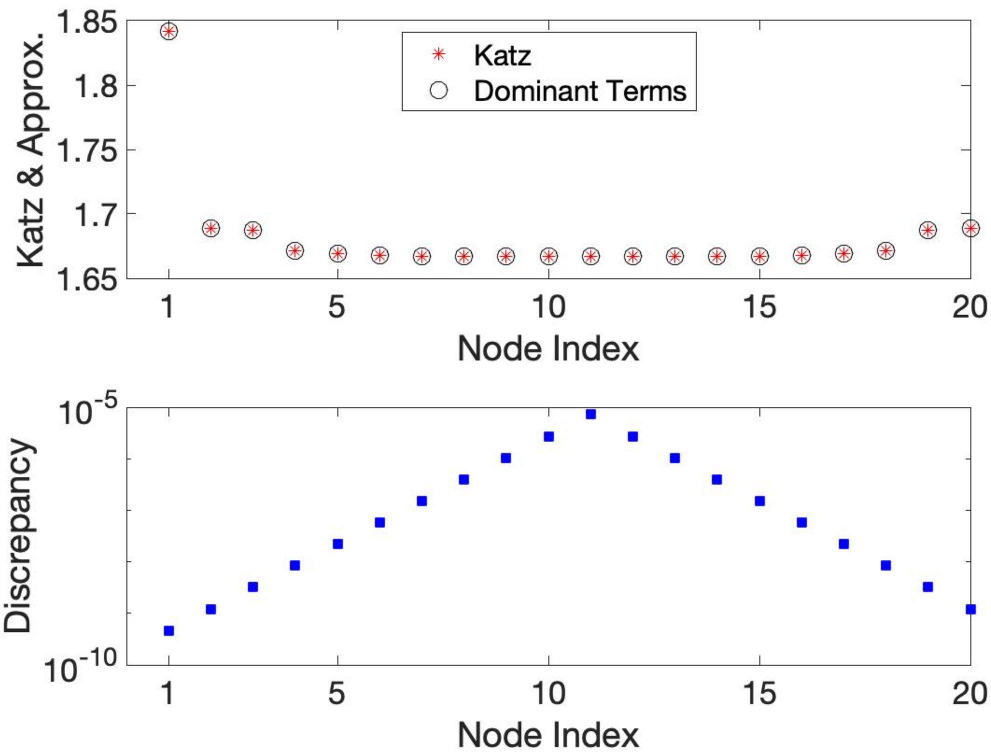



Figure 5.3: Upper picture: asterisks show components of Katz vector from (3.1) and circles show the approximation $b+h_1t_1^{p(i)}+h_3t_3^{p(i)}$ from (5.30). Here $b = 1/(1-4\alpha)$, $t_1$ and $t_3$ is defined in (5.29), $h_1$ and $h_3$ was found by solving (5.50). Lower picture: the discrepancy $x_i - b - h_1t_1^{p(i)} - h_3t_3^{p(i)}$ on a log scale. From Theorem 5.2, this quantity has the form $h_2t_2^{p(i)} + h_4t_4^{p(i)}$, and hence grows geometrically away from the shortcut node. However, it is uniformly $O(t_1^{N/2}) + O(|t_3|^{N/2})$ for a fixed $0 < t_1 < 1$ and $-1 < t_3 < 0$, and hence rapidly becomes negligible as the network size $N$ increases.

*where* $h_1 > h_3$ *with* $h_1 = O(1)$ *and* $h_3 = O(1)$.

*Proof.* Our proof proceeds in a similar manner to previous proofs. We begin by assuming that we have only spike centered at node 1 and use the anzatz

$$x_i = b + ht^{p(i)}. \tag{5.31}$$

The Katz system (3.1) essentially reduces to five scalar equations where we need to consider:

- General node $k$ where each neighbouring node differs by a factor of $t$.
- Node 1 which has a shortcut across the ring to node $L$.
- A node that neighbours node 1, i.e., node 2 or node $N$.
- The node(s) furthest away from node 1 periodically, i.e.,
  - node $q_1 = N/2 + 1$, if $N$ is even,
  - node $q_1 = (N + 1)/2$ or $q_1 = (N + 3)/2$, if $N$ is odd.
- A node that neighbours the furthest away node from node $N$ (the smallest ranking node(s)). If $N$ is even then node $q_1 - 1$ or node $q_1 + 1$. If $N$ is odd, and $q_1 = (N+1)/2$ then node $q_1 - 1$ or $q_1 + 2$ and similarly, if $q_1 = (N+3)/2$ then node $q_1 - 2$ or $q_1 + 1$.

For the purpose of this proof, when we consider the case of $N$ odd we will use $q_1 = (N + 1)/2$. When we have a choice of neighbouring nodes we will examine node 2 and $q_1 - 1$. Other choices are equally valid and will result in the same analysis.

At a general node $k$, corresponding to the $k$th row of the linear system, for $k \neq 1$, $k \neq 2$, $k \neq N$, $k \neq q_1$, $k \neq q_1 - 1$ and $k \neq q_1 + 1$ we have

$$-\alpha x_{k-2} - \alpha x_{k-1} + x_k - \alpha x_{k+1} - \alpha x_{k+2} = 1. \tag{5.32}$$

Inserting (5.31), this becomes

$$b(1-4\alpha) + ht^{p(k)}\left[1 - \alpha t^2 - \alpha t - \alpha/t - \alpha/t^2\right] = 1.$$

We can satisfy this equation independently of $k$ by setting $b = 1/(1-4\alpha)$ and choosing $t$ to be roots of the quartic polynomial $\alpha t^4 + \alpha t^3 - t^2 + \alpha t + \alpha$. By linearity of (5.32), we may extend (5.31) to include a linear combination involving all four roots, so our ansatz becomes

$$x_k = b + h_1 t_1^{p(k)} + h_2 t_2^{p(k)} + h_3 t_3^{p(k)} + h_4 t_4^{p(k)}. \tag{5.33}$$

The rates $t_1, t_2, t_3$ and $t_4$ are given by (5.29). The roots appear in reciprocal pairs such that $t_1 t_2 = 1$ and $t_3 t_4 = 1$. In addition we have $t_4 < -1 < t_3 < 0 < t_1 < 1 < t_2$ and given the form of the polynomial we note that

$$1 - \alpha t_j^2 - \alpha t_j = \alpha t_j^{-1} + \alpha t_j^{-2}. \tag{5.34}$$

The four remaining equations to satisfy from (3.1) arise at nodes 1, 2, $q_1$ and $q_1 - 1$, where we require

$$-\alpha x_{N-1} - \alpha x_N + x_1 - \alpha x_2 - \alpha x_3 - \alpha x_L = 1, \tag{5.35}$$

$$-\alpha x_N - \alpha x_1 + x_2 - \alpha x_3 - \alpha x_4 = 1, \tag{5.36}$$

$$\begin{cases} -\alpha x_{N/2-1} - \alpha x_{N/2} + x_{N/2+1} - \alpha x_{N/2+2} - \alpha x_{N/2+3} = 1, & \text{(5.37a)} \\ -\alpha x_{(N-3)/2} - \alpha x_{(N-1)/2} + x_{(N+1)/2} - \alpha x_{(N+3)/2} - \alpha x_{(N+5)/2} = 1, & \text{(5.37b)} \end{cases}$$

and

$$\begin{cases} -\alpha x_{N/2-2} - \alpha x_{N/2-1} + x_{N/2} - \alpha x_{N/2+1} - \alpha x_{N/2+2} = 1, & \text{(5.38a)} \\ -\alpha x_{(N-5)/2} - \alpha x_{(N-3)/2} + x_{(N-1)/2} - \alpha x_{(N+1)/2} - \alpha x_{(N+3)/2} = 1, & \text{(5.38b)} \end{cases}$$

respectively. Equations (5.37a) and (5.38a) are valid when $N$ is even and (5.37b) and (5.38b) are valid when $N$ is odd.

Inserting (5.33) into (5.35) – (5.38b), using (5.34) and $b(1-4\alpha)=1$, we have

$$\sum_{j=1}^{4} h_j \left[t_j^{-2} + t_j^{-1} - t_j - t_j^2 - t_j^{L-1}\right] = b, \tag{5.39}$$

$$\sum_{j=1}^{4} h_j t_j \left[t_j^{-2} - 1\right] = 0, \tag{5.40}$$

$$\begin{cases} \sum_{j=1}^{4} h_j t_j^{N/2} \left[t_j^2 + t_j - t_j^{-1} - t_j^{-2}\right] = 0, & \text{if } N \text{ is even} \quad (5.41a) \\ \sum_{j=1}^{4} h_j t_j^{(N-1)/2} \left[t_j^2 + t_j - 1 - t_j^{-1}\right] = 0, & \text{if } N \text{ is odd} \quad (5.41b) \end{cases}$$

and

$$\begin{cases} \sum_{j=1}^{4} h_j t_j^{N/2-1} \left[t_j^2 - 1\right] = 0, & \text{if } N \text{ is even} \quad (5.42a) \\ \sum_{j=1}^{4} h_j t_j^{(N-1)/2} \left[t_j - 1\right] = 0, & \text{if } N \text{ is odd.} \quad (5.42b) \end{cases}$$

We will now consider the cases of $N$ even and $N$ odd to show that (5.41a) – (5.42b) can be reduced to the same two conditions.

**Consider the case when the number of nodes, $N$, is even.**

Using the fact that $t_1t_2 = 1$ and $t_3t_4 = 1$ we can simplify (5.41a) to get

$$\left[t_2^2 + t_2 - t_1 - t_1^2\right]\left[-h_1t_1^N + h_2\right]t_1^{-N/2} + \left[t_4^2 + t_4 - t_3 - t_3^2\right]\left[-h_3t_3^N + h_4\right]t_3^{-N/2} = 0, \tag{5.43}$$

and simplify (5.42a) to get

$$\left[t_2 - t_1\right]\left[-h_1t_1^N + h_2\right]t_1^{-N/2} + \left[t_4 - t_3\right]\left[-h_3t_3^N + h_4\right]t_3^{-N/2} = 0. \tag{5.44}$$

In order to satisfy both (5.43) and (5.44) it follows that

$$\left[t_2^2 - t_1^2\right]\left[-h_1t_1^N + h_2\right]t_1^{-N/2} + \left[t_4^2 - t_3^2\right]\left[-h_3t_3^N + h_4\right]t_3^{-N/2} = 0, \tag{5.45}$$

which can only be true if

$$h_2 = h_1 t_1^N, \quad h_4 = h_3 t_3^N. \tag{5.46}$$

**Consider the case when the number of nodes, $N$, is odd.**

Using the fact that $t_1 t_2 = 1$ and $t_3 t_4 = 1$ we can simplify (5.41b) to get

$$\begin{aligned} h_1 t_1^{(N-1)/2}\left[t_1^2 + t_1 - t_2 - 1\right] + h_2 t_2^{(N-1)/2}\left[t_2^2 + t_2 - t_1 - 1\right] \\ + h_3 t_3^{(N-1)/2}\left[t_3^2 + t_3 - t_4 - 1\right] + h_4 t_4^{(N-1)/2}\left[t_4^2 + t_4 - t_3 - 1\right] = 0, \end{aligned} \tag{5.47}$$

and simplify (5.42b) to get

$$\begin{aligned} h_1 t_1^{(N-1)/2}\left[t_1 - 1\right] + h_2 t_2^{(N-1)/2}\left[t_2 - 1\right] \\ + h_3 t_3^{(N-1)/2}\left[t_3 - 1\right] + h_4 t_4^{(N-1)/2}\left[t_4 - 1\right] = 0. \end{aligned} \tag{5.48}$$

In order to satisfy both (5.47) and (5.48) it follows that

$$\left[t_2^{3/2} - t_1^{3/2}\right]\left[-h_1 t_1^N + h_2\right] t_1^{-N/2} + \left[t_4^{3/2} - t_3^{3/2}\right]\left[-h_3 t_3^N + h_4\right] t_3^{-N/2} = 0, \tag{5.49}$$

which can only be true if

$$h_2 = h_1 t_1^N, \quad h_4 = h_3 t_3^N. \tag{5.46}$$

**Continuation of general case: $N$ even or odd.**

We have shown that the conditions imposed at nodes $q_1$ and $q_1 - 1$ can be reduced to the same conditions irrespective of whether the number of nodes is even or odd. We may now use the relation (5.46) to eliminate $h_2$ and $h_4$ in (5.39) and (5.40), thus reducing the system to 2 equations in 2 unknowns. This may be written as

$$(B + \Delta B)\, v = r, \tag{5.50}$$

where

$$B = \begin{bmatrix} t_2^2 + t_2 - t_1 - t_1^2 & t_4^2 + t_4 - t_3 - t_3^2 \\ t_2 - t_1 & t_4 - t_3 \end{bmatrix}, \quad v = \begin{bmatrix} h_1 \\ h_3 \end{bmatrix}, \quad r = \begin{bmatrix} b \\ 0 \end{bmatrix},$$

and $\Delta B \in \mathbb{R}^{2\times 2}$ is such that $\|\Delta B\|_\infty = O(t_1^{\theta N/2}) + O(t_3^{\theta N/2})$. The determinant of $B$ is given by

$$\left[t_2^2 + t_2 - t_1 - t_1^2\right](t_4 - t_3) - \left[t_4^2 + t_4 - t_3 - t_3^2\right](t_2 - t_1). \tag{5.51}$$

Since $t_4 < -1 < t_3 < 0 < t_1 < 1 < t_2$ it follows that $t_4 - t_3 < 0$ and $t_2 - t_1 > 0$. Using (5.29) and noting that $x = \sqrt{4/\alpha + 9} \geq 3$ we have

$$\begin{aligned} \left[t_2^2 + t_2 - t_1 - t_1^2\right] &= \frac{1}{4}(x+1)\sqrt{y - 2x} > 0 \\ \left[t_4^2 + t_4 - t_3 - t_3^2\right] &= \frac{1}{4}(x-1)\sqrt{y + 2x} > 0. \end{aligned}$$

Therefore the first part of (5.51) is negative and the second part of (5.51) is positive. It then follows that the determinant is negative so $B$ is non-singular.

From (5.50), we have

$$\left(I + B^{-1}\Delta B\right) v = B^{-1} r.$$

The matrix $I + B^{-1}\Delta B$ is invertible for sufficiently large $N$, which establishes that (5.50) has a unique solution. We conclude that $x$ in (5.33) solves the Katz system (3.1). Moreover, since $(I + B^{-1}\Delta B)^{-1} = I + O(\Delta B)$, we see that $v = B^{-1}r + O(t_1^{\theta N/2}) + O(|t_3|^{\theta N/2})$, which leads to

$$\begin{bmatrix} h_1 \\ h_3 \end{bmatrix} \approx \frac{1}{\det(B)} \begin{bmatrix} b(t_4 - t_3) \\ -b(t_2 - t_1) \end{bmatrix}.$$

We know that $b = 1/(1 - 4\alpha) > 0$, $t_4 - t_3 < 0$, $t_2 - t_1 > 0$ and we have shown that $\det(B) < 0$. It then follows that $h_1 > 0$ and $h_3 > 0$. Using the relations (5.46) we also have that $h_2 > 0$ and since $t_3 < 0$, $h_4 > 0$ if $N$ is even and $h_4 < 0$ if $N$ is odd. Since $t_4 - t_3 = -\frac{1}{2}\sqrt{y + 2x}$ and $t_2 - t_1 = \frac{1}{2}\sqrt{y - 2x}$ it follows that $|t_4 - t_3| > |t_2 - t_1|$ and therefore $h_1 > h_3$.

As with the approach in previous theorems, we can use (5.46) to deduce that $h_2 = O(t_1^N)$ and $h_4 = O(|t_3|^N)$. It follows that $h_2 t_2^{p(k)} = h_1 t_1^{N-p(k)}$ and $h_4 t_4^{p(k)} = h_3 t_3^{N-p(k)}$ is largest when $k = N/2 + 1$ so

$$\begin{aligned} x_i &= b + h_1 t_1^{p(i)} + h_2 t_2^{p(i)} + h_3 t_3^{p(i)} + h_4 t_4^{p(i)} \\ &= b + h_1 t_1^{p(i)} + h_3 t_3^{p(i)} + O(t_1^{N/2}) + O(|t_3|^{N/2}). \end{aligned}$$

□

Since $t_4 < -1 < t_3 < 0$ we observe geometric decay and growth terms that change sign depending on whether the node is even or odd. We can explain the rationale of why this may be the case. In effect the $h_1 t_1^{p(i)}$ and $h_2 t_2^{p(i)}$ terms can be thought of as representing the 2-nearest neighbours ring and the $h_3 t_3^{p(i)}$ and $h_4 t_4^{p(i)}$ can be thought of as representing the next nearest neighbour ring.

Consider a ring in which each node did not connect to its nearest neighbour but was instead connected to its second nearest neighbour. If we have an even number of nodes, $N$ even, then the ring would not be considered well connected as the network could be split into two disjoint sets – one with the odd nodes and one with the even nodes. Should we add a shortcut across this ring from node 1 to node $L$ then we would have a spike at node 1, geometric decay around the odd nodes and $N/2 + 1$ would be the least ranked odd node. The even nodes would all be ranked equal with the baseline centrality score. Considering the odd nodes we have geometric decay at node 1 and geometric growth from node 1 that peaks at node $N/2 + 1$ (an odd node).

Should we have an odd number of nodes, $N$ odd, then the network cannot be split into two disjoint sets as the two odd nodes 1 and $N$ are neighbours. In this case information can flow directly via the edges with periodic distance 2 through both the odd and even nodes.

Now when $N$ is even we have $h_3 t_3^{p(i)}$ and $h_4 t_4^{p(i)}$ which are positive for odd nodes and negative for even nodes. We are in effect, adding the geometric decay and growth terms to the odd nodes to account for the shortcut that the odd nodes can access and subtracting them from the even nodes.

When $N$ is odd, $h_3 t_3^{p(i)}$ is as before - positive for odd nodes and negative for even nodes. However, $h_4 t_4^{p(i)}$ is negative for odd nodes and positive for even nodes. In this instance, we are adding the decay term and subtracting the growth term from the odd nodes.

We hypothesise that in the general case of a $2n$-nearest neighbour periodic ring that connects to $n$-nearest nodes on each side with a shortcut across the ring from node 1 to node $L$ that the solution of the Katz system (3.1) has the form

$$x_i = b + \sum_{j=1}^{2n} h_j t_j^{p(i)} \tag{5.52}$$

where $b = 1/(1 - 2\alpha n)$ and $t_j$ are the roots of the palindromic polynomial $\alpha t^{2n} + \cdots + \alpha t^{n+1} - t^n + \alpha t^{n-1} + \cdots + \alpha$.

Thus far we have only considered the instances where the roots of the palindromic polynomial are real. In considering higher order polynomials, for example, the degree 6 polynomial

$$\alpha t^6 + \alpha t^5 + \alpha t^4 - t^3 + \alpha t^2 + \alpha t + \alpha, \tag{5.53}$$

resulting from each node of the ring connecting to 3 nearest neighbours on each side we encounter complex $t$. We know that $\rho(A) \leq ||A||_\infty = 1/7$ for such an adjacency matrix so it follows that $1/7 \leq 1/\rho(A)$. For example, solving for the roots of (5.53) with $\alpha = 0.1 < 1/\rho(A)$ we find

$$\begin{aligned} t_1 &\approx 0.6081, \quad t_3 \approx -0.2380 - 0.3388i, \quad t_5 \approx -0.2380 + 0.3388i, \\ t_2 &\approx 1.6446, \quad t_4 \approx -1.3884 + 1.9765i, \quad t_6 \approx -1.3884 - 1.9765i. \end{aligned}$$

We know that since the polynomial is palindromic the roots appear as reciprocal pairs and as demonstrated above as complex conjugate pairs.

In Chapter 2 we discussed properties of palindromic polynomials with real coefficients noting that if complex roots exist then we encounter them in groups of four complex roots: $z, \overline{z}, z^{-1}$ and $\overline{z}^{-1}$. We note here that for the above example the complex roots do not lie on the unit circle and thus we have four distinct roots.

We conjecture that in the case of complex $t$, the combination in (5.52), cancels the imaginary terms by having some $h_{j_1} = \overline{h}_{j_2}$. This would however require further study to determine.

In the case of this network class where the sub-diagonals and super-diagonals of the adjacency matrix representing the circulant ring are equal (same number of connections on each side of the node) the general node equation will always give rise to a palindromic polynomial. Under our ansatz, this palindromic polynomial with roots appearing in reciprocal pairs will determine the rate of both geometric decay and growth on moving away from the node with the shortcut. We rule out the possibility of unit roots as this could not give rise to geometric decay or growth. We therefore conjecture it always to be the case that we will have $n$ roots with $|t| < 1$ and $n$ roots with $|t| > 1$.

# Chapter 6

# Example networks

The results from Theorems 4.1, 4.2 and 5.1 give exact characterizations of the centrality vectors, and the presence of both geometrically decaying and *geometrically increasing* components is not intuitively obvious. However, the qualitative nature of the solution, with most weight given to the node with the highest out-degree and with overall centrality decaying according to periodic distance from this node, is no surprise. In this Chapter, we show that the type of analysis developed here can be applied to more general networks where the results are not predictable. To our knowledge, this is the first example to capture analytically (rather than experimentally) a change in node centrality ranking as the Katz parameter is varied. To isolate the key ideas, we have chosen simple network structures, but we note that the same approach can be applied in more general settings.

In this Chapter we study a class of networks built from $R$ identical undirected, periodic, nearest neighbour $m$-node rings. Nodes 1 to $m$, $m+1$ to $2m$, ..., $(R-1)m+1$ to $Rm$ make up rings $1, 2, \ldots, R$ respectively. We also introduce a 'hub' node $N := Rm + 1$ that connects to local node 1 in each ring. Within the following sections we examine the networks generated when (a) we add a directed or undirected shortcut across each of the $R$ rings and (b) the hub node edge(s) are directed or undirected. For examples of the networks that we study please refer to Figures 6.1, 6.5, 6.9 and 6.13.

We note that as we increase Katz parameter, $\alpha$, from 0 to $1/\rho(A)$ we interpolate between degree and eigenvector centrality [21]. We will therefore focus this Chapter on considering the Katz centrality. By symmetry, we only need to consider centrality for nodes in the first ring, that is, nodes 1 to $m$, and for the additional node, $N$. We use an ansatz from the respective theorem, but rather than deriving an exact solution, we focus on the dominant term and proceed heuristically. We then follow up the analysis with computational results.

In this class of networks, the edges possessed by node $N$ connect to nodes that are themselves well-connected and can propogate more walks around the network than other nodes on the ring. This suggests that as $\alpha$ is increased, and hence longer walks become more relevant, Katz centrality may give more relative weight to node $N$. Our aim is to quantify this effect.

All figures which relate to the following sections have been placed at the end of their respective section.

## 6.1 $R$-rings with a directed shortcut

In this section we consider the case of each ring in the network having a directed shortcut across the ring. As each ring is identical, we denote this using ring 1 as an edge from node 1 to node $L$ where $1 \ll L \leq m/2+1$. In considering this ring with a directed shortcut we would expect node 1 to attain the maximum score and define the node with the minimal score as node $\delta$. That is, node 1 will be ranked first and node $\delta$ will be ranked last. Should $L = m/2+1$ then node $\delta = L$. We explore how the influence of the Katz parameter $\alpha$ effects the ranking of the central node $N$.

### 6.1.1 Connection to a directed hub

This subsection considers the case of each of the $R$ rings connecting to the central hub node with an incoming directed edge from central hub node $N$ to each of the local node 1s. Figure 6.1 illustrates an example network with $R = 3$ rings where each ring contains the same identical directed shortcut from local node 1 to local node 6 (i.e. $1 \mapsto 6, 11 \mapsto 16, 21 \mapsto 26$). Each local node 1 is connected to the central node 31 via an incoming directed edge (i.e. $31 \mapsto 1, 31 \mapsto 11, 31 \mapsto 21$).

For the general $R$-ring network, the Katz system (3.1) reduces to

$$x_j - \alpha\,(x_{j-1} + x_{j+1}) = 1, \qquad \text{for } 2 \le j \le m, \tag{6.1}$$

$$x_1 - \alpha\,(x_2 + x_m + x_L) = 1, \tag{6.2}$$

$$x_N - R\alpha x_1 = 1. \tag{6.3}$$

Inserting the ansatz (4.8) we find that, as in the proof of Theorem 4.1, the general equation (6.1) is solved with $b = 1/(1-2\alpha)$ and $t = t_1$. Using $x_L \approx b$ and solving (6.2) for $h$, we arrive at

$$h \approx \frac{\alpha}{(1-2\alpha)\sqrt{1-4\alpha^2}}. \tag{6.4}$$

Requiring $h > 0$ in (6.4) places the restriction that $0 < \alpha < \widehat{\alpha}$ where $\widehat{\alpha} = 1/2$. We therefore conjecture that the network adjacency matrix has a spectral radius that approaches 2 as $N \to \infty$. Table 6.1 gives an indication of the discrepancy between the spectral radius of the adjacency matrix ($\rho(A)$) and the reciprocal of the upper bound placed on the $\alpha$ domain ($1/\widehat{\alpha}$) such that $h > 0$.

| $R$ | $\widehat{\alpha}$ | $1/\widehat{\alpha}$ | $\lvert\rho(A) - 1/\widehat{\alpha}\rvert$ | | |
|---|---|---|---|---|---|
| | | | $m = 10^2$ | $m = 10^4$ | $m = 10^6$ |
| 1 | 1/2 | 2 | $\approx 3.3 \times 10^{-3}$ | $\approx 4.3 \times 10^{-4}$ | $\approx 6.2 \times 10^{-6}$ |
| 2 | 1/2 | 2 | $\approx 3.3 \times 10^{-3}$ | $\approx 4.3 \times 10^{-4}$ | $\approx 6.2 \times 10^{-6}$ |
| 3 | 1/2 | 2 | $\approx 3.3 \times 10^{-3}$ | $\approx 4.3 \times 10^{-4}$ | $\approx 6.2 \times 10^{-6}$ |
| 4 | 1/2 | 2 | $\approx 3.3 \times 10^{-3}$ | $\approx 4.2 \times 10^{-4}$ | $\approx 6.2 \times 10^{-6}$ |

Table 6.1: *$R$ rings with directed shortcut and directed hub*: Comparison of $\rho(A)$ and $1/\widehat{\alpha}$ where $\widehat{\alpha}$ is the upper bound of the domain defined by the heuristic analysis. [Due to convergence issues in using MATLAB's `eigs`, $\rho(A)$ was computed using the power method.]

We know that $x_1 = b + h$ and from (6.3) that $x_N = 1 + R\alpha(b + h)$. We now turn our attention to when $x_N > x_1$, i.e.,

$$1 + (R\alpha - 1)(b + h) > 0.$$

Using the expression for $h$ in (6.4) this gives

$$1 + (R\alpha - 1)\left[\frac{1}{1-2\alpha} + \frac{\alpha}{(1-2\alpha)\sqrt{1-4\alpha^2}}\right] > 0$$

which rearranges to

$$(1-2\alpha)\sqrt{1-4\alpha^2} + (R\alpha - 1)[\sqrt{1-4\alpha^2} + \alpha] > 0.$$

We arrive at

$$z_1 := (R-2)\sqrt{1-4\alpha^2} + R\alpha - 1 > 0. \tag{6.5}$$

From Figure 6.2 we observe that $z_1 > 0$ for $R \geq 3$ and $z_1 < 0$ for $R < 3$ for $0 < \alpha < 1/2$ which covers our whole $\alpha$ domain for which the Katz system is valid. We can therefore conclude that $x_N > x_1$ when we have 3 or more rings in our network and $x_1 > x_N$ when we have fewer than 3 rings.

In the case of having fewer than 3 rings, a natural extension is to evaluate how $x_N$ compares with the other nodes in the ring. To do so we first compare with node $\delta$ (the minimum scored node). Using $x_\delta \approx b$ we find $x_N > x_\delta$ when

$$1 + R\alpha(b + h) > b.$$

Using the expression for $h$ in (6.4) this gives

$$1 + R\alpha\left[\frac{1}{1-2\alpha} + \frac{\alpha}{(1-2\alpha)\sqrt{1-4\alpha^2}}\right] > \frac{1}{1-2\alpha}$$

which rearranges to

$$(1-2\alpha)\sqrt{1-4\alpha^2} + R\alpha[\sqrt{1-4\alpha^2} + \alpha] > \sqrt{1-4\alpha^2}.$$

We arrive at

$$R\alpha + (R-2)\sqrt{1-4\alpha^2} > 0. \tag{6.6}$$

It is clear that in the case when $R = 2$, (6.6) is satisfied for $0 < \alpha < 1/2$ indicating that $x_N > x_\delta$.

In the case when $R = 1$, we require $\alpha > \sqrt{1-4\alpha^2}$ indicating that there is a critical value $\alpha_1^* = 1/\sqrt{5}$ for which the nodal rankings change. We therefore have, in the case of 1 ring, $x_\delta > x_N$ when $0 < \alpha < \alpha_1^*$ and $x_N > x_\delta$ when $\alpha_1^* < \alpha < 1/2$.

A summary of findings thus far can be found in Table 6.2.

| R | Degree | Katz | Eigenvector |
|---|---|---|---|
| 1 | $x_1 > x_\delta > x_N$ | $x_1 > x_\delta > x_N \ \forall \alpha \in (0, \alpha_1^*)$<br>$x_1 > x_N > x_\delta \ \forall \alpha \in (\alpha_1^*, \widehat{\alpha})$ | $x_1 > x_N > x_\delta$ |
| 2 | $x_1 > x_N = x_\delta$ | $x_1 > x_N > x_\delta \ \forall \alpha \in (0, \widehat{\alpha})$ | $x_1 > x_N > x_\delta$ |
| 3 | $x_N = x_1 > x_\delta$ | $x_N > x_1 > x_\delta \ \forall \alpha \in (0, \widehat{\alpha})$ | $x_N > x_1 > x_\delta$ |
| $\geq 4$ | $x_N > x_1 > x_\delta$ | $x_N > x_1 > x_\delta \ \forall \alpha \in (0, \widehat{\alpha})$ | $x_N > x_1 > x_\delta$ |

Table 6.2: *R rings with directed shortcut and directed hub*: Katz centrality for $x_1, x_\delta$ and $x_N$ from the heuristic analysis compared with degree ($\alpha \to 0$) and eigenvector ($\alpha \to 1/\rho(A)$) centrality

Figure 6.3 shows the ratio of $x_1$ and $x_\delta$ to $x_N$ as $\alpha$ is varied for $1 \leq R \leq 4$. To do this, we used $m = 1{,}000$ with $L = 501$ and solved the Katz system (3.1) directly. By examining these figures we can see that the results from our heuristic analysis, presented consisely in Table 6.2, agree with the solution of the Katz system. In particular, Figure 6.3a shows the change in ranking of nodes $\delta$ and $N$ on passing the threshold $\alpha_1^*$.

In the case when node $N$ is neither ranked first nor last, i.e. when $x_1 > x_N > x_\delta$, we can quantify where node $N$ is ranked in relation to nodes in the ring by considering where $x_k > x_N$. This condition can be reduced to

$$b + ht_1^{p(k)} > 1 + R\alpha(b+h).$$

Using the expression for $h$ in (6.4) this gives

$$\frac{\alpha}{(1-2\alpha)\sqrt{1-4\alpha^2}}\left[t_1^{p(k)} - R\alpha\right] > 1 + \frac{R\alpha - 1}{1-2\alpha}$$

which rearranges to

$$\alpha\left[t_1^{p(k)} - R\alpha\right] > (1-2\alpha)\sqrt{1-4\alpha^2} + (R\alpha - 1)\sqrt{1-4\alpha^2}.$$

We arrive at

$$t_1^{p(k)} > R\alpha + (R-2)\sqrt{1-4\alpha^2}.$$

As expected we find that the ranking of the central node is dependent upon the choice of $\alpha$ and that, in particular, $x_k > x_N$ for $p(k) < p(k)_{\max}$ where

$$p(k)_{\max} = \begin{cases} \dfrac{\log(\alpha - \sqrt{1-4\alpha^2})}{\log(1-\sqrt{1-4\alpha^2}) - \log 2\alpha}, & \text{for } R = 1 \text{ and } \alpha_1^* < \alpha < \widehat{\alpha} \quad (6.7\text{a}) \\ \dfrac{\log 2\alpha}{\log(1-\sqrt{1-4\alpha^2}) - \log 2\alpha}, & \text{for } R = 2 \text{ and } 0 < \alpha < \widehat{\alpha}. \quad (6.7\text{b}) \end{cases}$$

The functions (6.7a) and (6.7b) are shown in Figure 6.4. In the case of only having 1 ring in the network, when $\alpha_1^* < \alpha < 1/2$, it follows that nodes with $p(k) < 5$ will always be deemed more central than node $N$. It is interesting to see that as $\alpha$ increases in this range the periodic distance decreases, indicating that node $N$ is becoming more important, before reaching a turning point and increasing at a quicker rate than it decreased. Figure 6.4b demonstrates that in the case of 2 rings although $x_1 > x_N > x_\delta$, since $p(k)_{\max} < 1$, node $N$ would be ranked between nodes 1 and 2 (or equally $m$), i.e. $x_N > x_2 = x_m$.

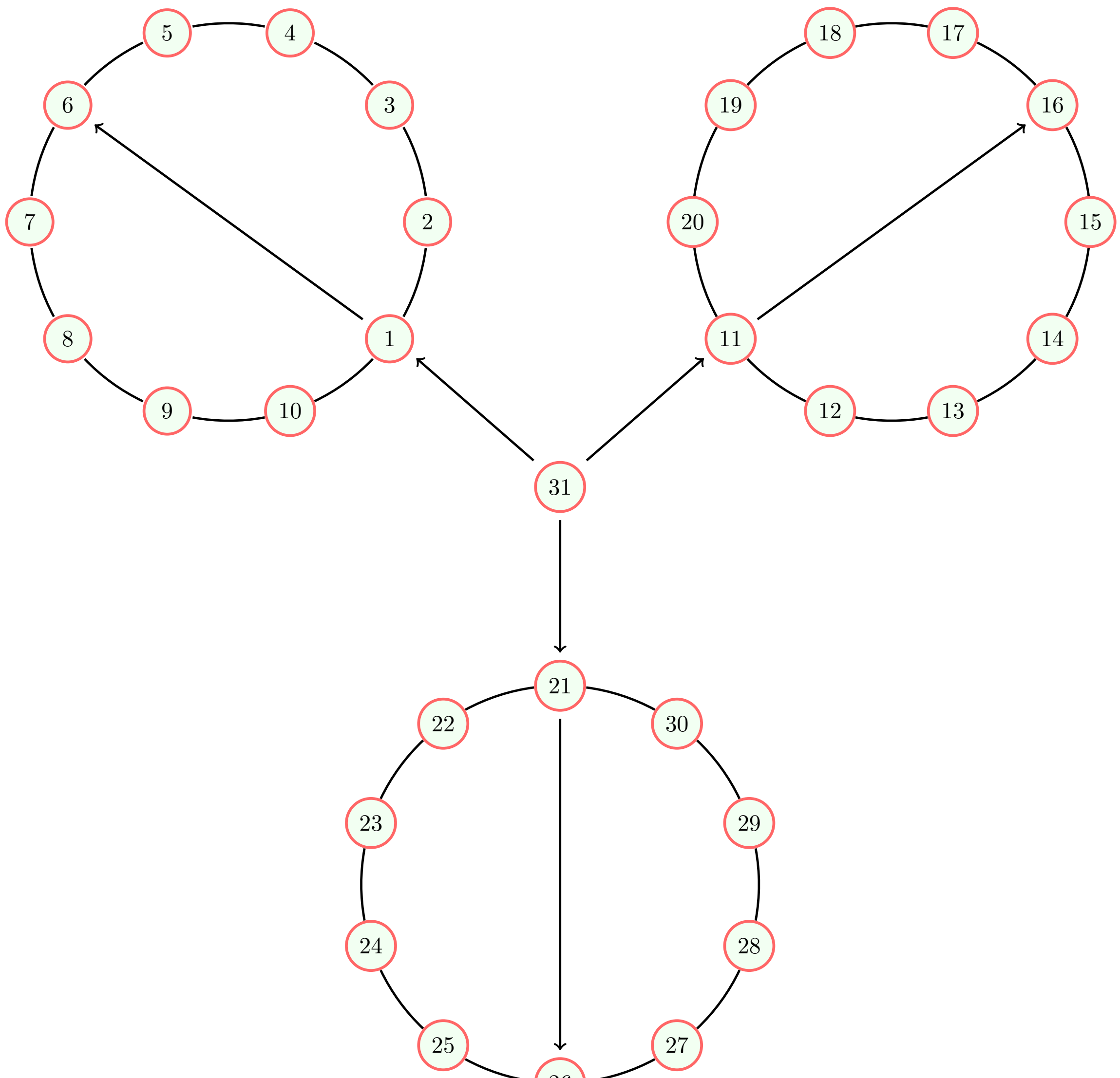
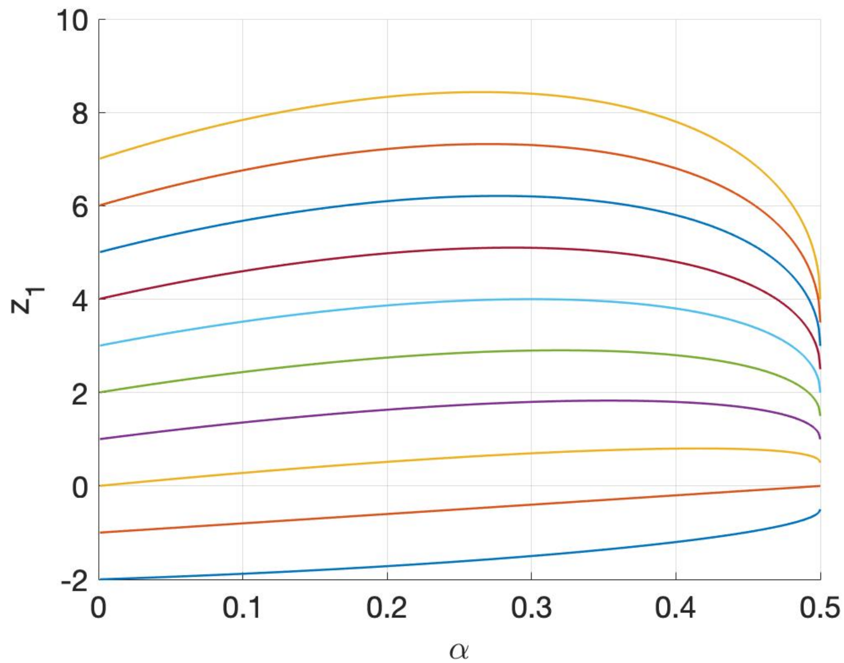
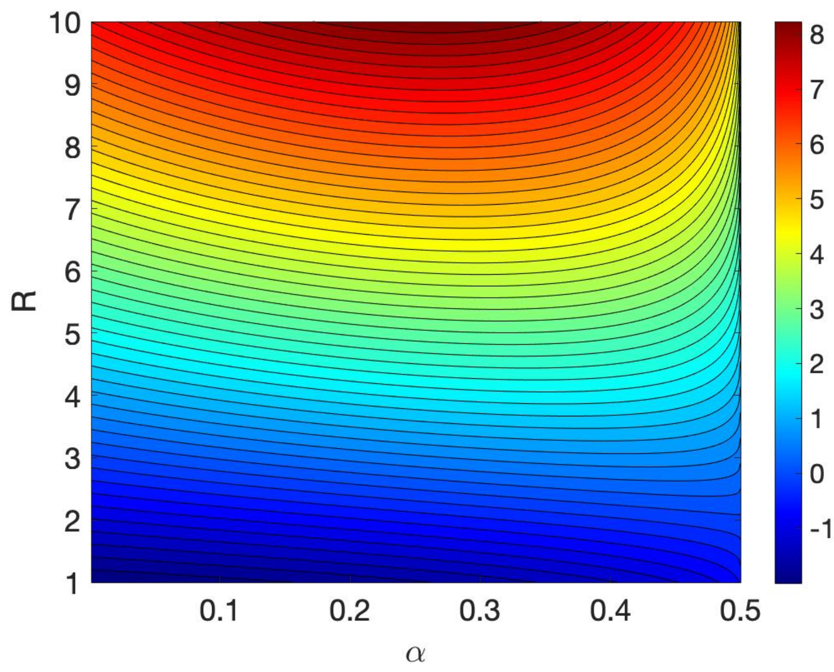



Figure 6.1: *R rings with directed shortcut and directed hub*: Illustration with directed shortcut and directed hub ($R = 3$, $m = 10$ and $L = 6$).

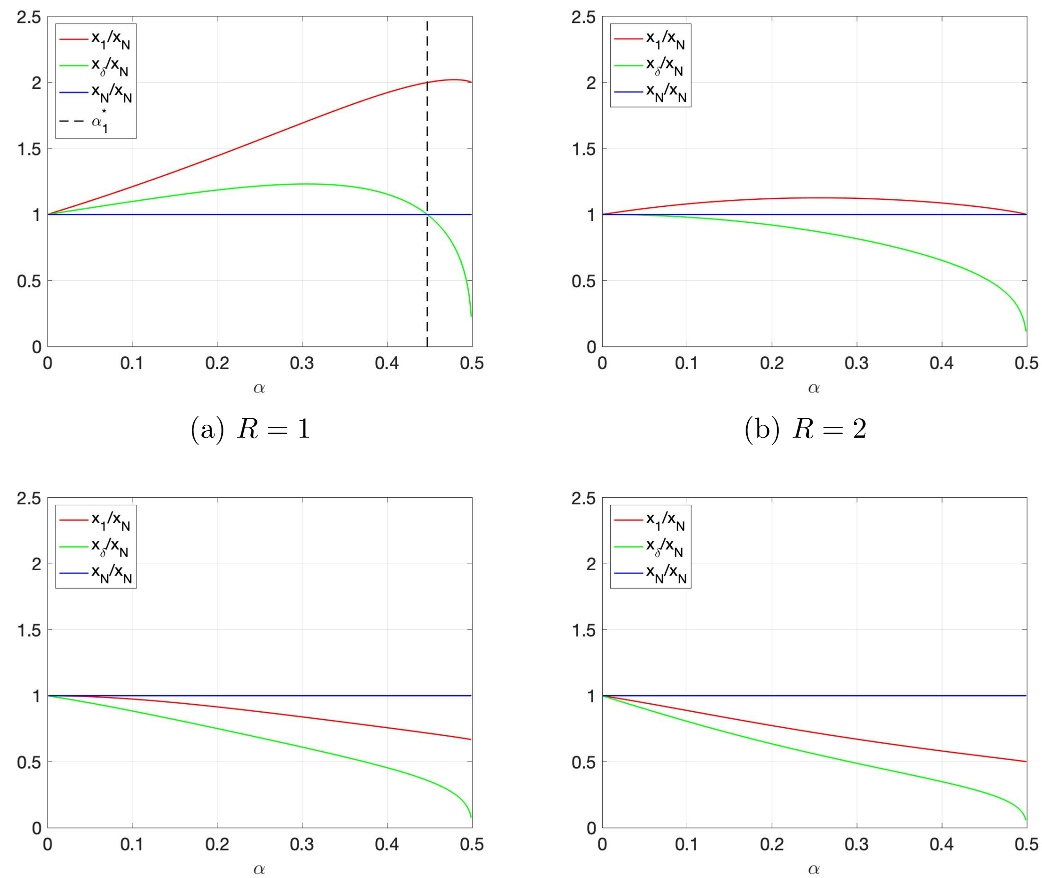



(a) Fixed $R$ ($R$=seq$(1, 10)$). $R = 1$ – bottom curve.




(b) Contour plot of $z_1$.

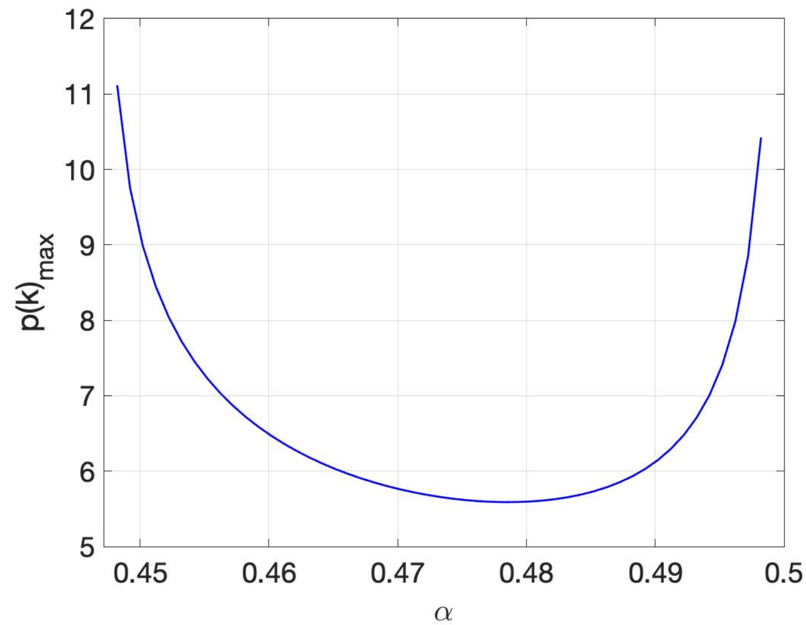
Figure 6.2: *R rings with directed shortcut and directed hub*: Relationship between number of rings $R$, Katz parameter $\alpha$ and $z_1$ in (6.5).



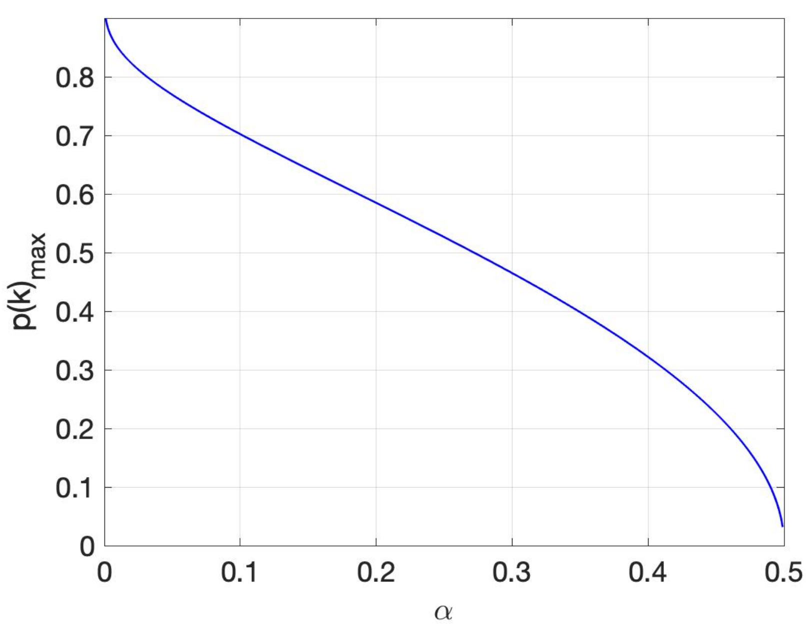

(a) $R = 1$ (b) $R = 2$

(c) $R = 3$ (d) $R = 4$

Figure 6.3: *R rings with directed shortcut and directed hub*: Ratio of $x_1$ and $x_\delta$ to $x_N$ for the Katz centrality. We used $m = 1,000$ so $N = 1000R + 1$ with $L = 501$.




(a) $R = 1$




(b) $R = 2$

Figure 6.4: *R rings with directed shortcut and directed hub*: $p(k)_{\max}$ indicating the maximum periodic distance for which $x_k > x_N$

### 6.1.2 Connection to an undirected hub

This subsection considers the case of each of the $R$ rings connecting to the central hub node with an undirected edge from central hub node $N$ to each of the local node 1s. Figure 6.5 illustrates an example network with $R = 3$ rings where each ring contains the same identical directed shortcut from local node 1 to local node 6 (i.e. $1 \mapsto 6, 11 \mapsto 16, 21 \mapsto 26$). Each local node 1 is connected to the central node 31 via an undirected edge (i.e. $1 \mapsto 31, 31 \mapsto 1, 11 \mapsto 31, 31 \mapsto 11, 21 \mapsto 31, 31 \mapsto 21$).

For the general $R$-ring network, the Katz system (3.1) reduces to

$$x_j - \alpha\,(x_{j-1} + x_{j+1}) = 1, \quad \text{for } 2 \le j \le m, \tag{6.8}$$

$$x_1 - \alpha\,(x_2 + x_m + x_L + x_N) = 1, \tag{6.9}$$

$$x_N - R\alpha x_1 = 1. \tag{6.10}$$

Inserting the ansatz (4.8) we find that, as in the proof of Theorem 4.1, the general equation (6.8) is solved with $b = 1/(1-2\alpha)$ and $t = t_1$. Using $x_L \approx b$ and solving (6.9) and (6.10) for $h$, we arrive at

$$h \approx \frac{\alpha\,(2 + (R-2)\alpha)}{(1-2\alpha)(\sqrt{1-4\alpha^2} - R\alpha^2)}. \tag{6.11}$$

Requiring $h > 0$ in (6.11) places the restriction that $\sqrt{1-4\alpha^2} - R\alpha^2 > 0$. Solving for the roots of the quartic polynomial $R^2\alpha^4 + 4\alpha^2 - 1$ we have

$$\alpha^2 = \frac{-2 \pm \sqrt{R^2+4}}{R^2}.$$

Since we require $\alpha \in \mathbb{R}^+$ it follows that $0 < \alpha < \widehat{\alpha}$ where

$$\widehat{\alpha} = \frac{\sqrt{-2 + \sqrt{R^2+4}}}{R}. \tag{6.12}$$

We therefore conjecture that the network adjacency matrix has a spectral radius that approaches $1/\widehat{\alpha}$ as $N \to \infty$ and hence the Katz system is valid for $0 < \alpha < \widehat{\alpha}$. Table 6.3 gives an indication of the discrepancy between the spectral radius of the adjacency matrix ($\rho(A)$) and the reciprocal of the upper bound placed on the $\alpha$ domain ($1/\widehat{\alpha}$) such that $h > 0$.

| $R$ | $\widehat{\alpha}$ | $1/\widehat{\alpha}$ | $\lvert\rho(A) - 1/\widehat{\alpha}\rvert$ | | |
|---|---|---|---|---|---|
| | | | $m = 10^2$ | $m = 10^4$ | $m = 10^6$ |
| 1 | $\approx 0.4859$ | $\approx 2.0582$ | $\approx 2.7 \times 10^{-6}$ | $\approx 4.0 \times 10^{-14}$ | $\approx 2.5 \times 10^{-14}$ |
| 2 | $\approx 0.4551$ | $\approx 2.1974$ | $\approx 1.9 \times 10^{-10}$ | $\approx 2.4 \times 10^{-14}$ | $\approx 2.4 \times 10^{-14}$ |
| 3 | $\approx 0.4224$ | $\approx 2.3676$ | $\approx 8.3 \times 10^{-14}$ | $\approx 1.4 \times 10^{-14}$ | $\approx 1.5 \times 10^{-14}$ |
| 4 | $\approx 0.3931$ | $\approx 2.5440$ | $\approx 8.9 \times 10^{-15}$ | $\approx 1.7 \times 10^{-14}$ | $\approx 1.4 \times 10^{-14}$ |

Table 6.3: *R rings with directed shortcut and undirected hub*: Comparison of $\rho(A)$ and $1/\widehat{\alpha}$ where $\widehat{\alpha}$ is the upper bound of the domain defined by the heuristic analysis. [$\rho(A)$ was computed using MATLAB's `eigs`.]

We know that $x_1 = b + h$ and from (6.10) that $x_N = 1 + R\alpha(b+h)$. We now turn our attention to when $x_N > x_1$, i.e.

$$1 + (R\alpha - 1)(b + h) > 0.$$

Using the expression for $h$ in (6.11) this gives

$$1 + (R\alpha - 1)\left[\frac{1}{1-2\alpha} + \frac{\alpha(2 + (R-2)\alpha)}{(1-2\alpha)(\sqrt{1-4\alpha^2} - R\alpha^2)}\right] > 0$$

which rearranges to

$$(R-2)(\sqrt{1-4\alpha^2} - R\alpha^2) + (R\alpha - 1)(2 + (R-2)\alpha) > 0.$$

We arrive at

$$z_2 := (R-2)\sqrt{1-4\alpha^2} + (R+2)\alpha - 2 > 0. \tag{6.13}$$

From Figure 6.6 we observe that $z_2 < 0$ for $R \leq 2$ and $z_2 > 0$ for $R \geq 4$ over the domain for which the Katz system is valid. We can therefore conclude that $x_N > x_1$ when we have 4 or more rings in our network and $x_1 > x_N$ when we have 2 or less rings.

In the case of having $R = 3$ rings, we require $\sqrt{1-4\alpha^2} + 5\alpha - 2 > 0$ and so solving for the roots of the quadratic $29\alpha^2 - 20\alpha + 3$ we observe that $z_2$ changes sign on crossing a pole $\alpha_2^* = (10 - \sqrt{13})/29$. We conjecture that this is the threshold value beyond which node $N$ is regarded as more central than node 1. The other pole of the quadratic, $(10+\sqrt{13})/29$, is outwith the domain of $\alpha$, i.e. $\widehat{\alpha} = \sqrt{-2+\sqrt{13}}/3 < (10+\sqrt{13})/29$.

In the case when node $N$ is not regarded as more central than node 1, a natural extension is to see how $x_N$ compares with the other nodes in the ring. To do so we first compare with node $\delta$ (the minimum scored node). Again, using $x_\delta \approx b$ we find $x_N > x_\delta$ when

$$1 + R\alpha(b+h) > b.$$

Using the expression for $h$ in (6.11) this gives

$$1 + R\alpha\left[\frac{1}{1-2\alpha} + \frac{\alpha\,(2+(R-2)\alpha)}{(1-2\alpha)(\sqrt{1-4\alpha^2}-R\alpha^2)}\right] > \frac{1}{1-2\alpha}$$

which rearranges to

$$(R-2)(\sqrt{1-4\alpha^2}-R\alpha^2) + R(2\alpha+(R-2)\alpha^2) > 0.$$

We arrive at

$$2R\alpha + (R-2)\sqrt{1-4\alpha^2} > 0. \tag{6.14}$$

It is clear to see that in the case when $R \geq 2$, (6.14) is satisfied for $0 < \alpha < \widehat{\alpha}$ meaning that $x_N > x_\delta$.

In the case when $R = 1$, we require $2\alpha - \sqrt{1-4\alpha^2} > 0$ indicating a critical value $\alpha_3^* = 1/\sqrt{8}$ where (6.14) changes sign. We can therefore only satisfy (6.14) for $\alpha_3^* < \alpha < \widehat{\alpha}$, meaning that $x_\delta > x_N$ when $0 < \alpha < \alpha_3^*$.

A summary of findings can be found in Table 6.4.

Figure 6.7 shows the ratio of nodes 1 and $\delta$ to node $N$ as $\alpha$ is varied for $1 \leq R \leq 4$. To do this, we used $m = 1,000$ with $L = 501$ and solved the Katz system (3.1) directly. By examining these figures we can see that the results from our heuristic analysis, presented in Table 6.4, agree with the solution of the Katz system. In particular, Figure 6.7a shows the change in ranking of nodes $\delta$ and $N$ on passing the threshold $\alpha_3^*$ and Figure 6.7c shows the change in ranking of nodes 1 and $N$ on passing the threshold $\alpha_2^*$.

| R | Degree | Katz | Eigenvector |
|---|---|---|---|
| 1 | $x_1 > x_\delta > x_N$ | $x_1 > x_\delta > x_N \ \forall\alpha \in (0, \alpha_3^*)$<br>$x_1 > x_N > x_\delta \ \forall\alpha \in (\alpha_3^*, \widehat{\alpha})$ | $x_1 > x_N > x_\delta$ |
| 2 | $x_1 > x_N = x_\delta$ | $x_1 > x_N > x_\delta \ \forall\alpha \in (0, \widehat{\alpha})$ | $x_1 > x_N > x_\delta$ |
| 3 | $x_1 > x_N > x_\delta$ | $x_1 > x_N > x_\delta \ \forall\alpha \in (0, \alpha_2^*)$<br>$x_N > x_1 > x_\delta \ \forall\alpha \in (\alpha_2^*, \widehat{\alpha})$ | $x_N > x_1 > x_\delta$ |
| $\geq 4$ | $x_N \geq x_1 > x_\delta$ | $x_N > x_1 > x_\delta \ \forall\alpha \in (0, \widehat{\alpha})$ | $x_N > x_1 > x_\delta$ |

Table 6.4: *R rings with directed shortcut and undirected hub*: Katz centrality for $x_1, x_\delta$ and $x_N$ from the heuristic analysis compared with degree ($\alpha \to 0$) and eigenvector ($\alpha \to 1/\rho(A)$) centrality

In the case when node $N$ is neither ranked first nor last, i.e. $x_1 > x_N > x_\delta$, we can quantify where node $N$ is ranked in relation to nodes in the ring by considering $x_k > x_N$. This condition can be reduced to

$$b + ht_1^{p(k)} > 1 + R\alpha(b + h).$$

Using the expression for $h$ in (6.11) this gives

$$\frac{\alpha\,(2 + (R-2)\alpha)}{(1-2\alpha)(\sqrt{1-4\alpha^2} - R\alpha^2)}\left[t_1^{p(k)} - R\alpha\right] > 1 + \frac{R\alpha - 1}{1 - 2\alpha}$$

which rearranges to

$$(2 + (R-2)\alpha)\left[t_1^{p(k)} - R\alpha\right] > (R-2)(\sqrt{1-4\alpha^2} - R\alpha^2).$$

We arrive at

$$t_1^{p(k)} > \frac{2R\alpha + (R-2)\sqrt{1-4\alpha^2}}{2 + (R-2)\alpha},$$

where we find that $p(k) < p(k)_{\max}$ for

$$p(k)_{\max} = \begin{cases} \dfrac{\log\,(2\alpha - \sqrt{1-4\alpha^2}) - \log(2-\alpha)}{\log\,(1 - \sqrt{1-4\alpha^2}) - \log 2\alpha}, & \text{for } R = 1 \text{ and } \alpha_3^* < \alpha < \widehat{\alpha} \quad (6.15a) \\ \dfrac{\log 2\alpha}{\log\,(1 - \sqrt{1-4\alpha^2}) - \log 2\alpha}, & \text{for } R = 2. \quad (6.15b) \\ \dfrac{\log\,(6\alpha + \sqrt{1-4\alpha^2}) - \log(2+\alpha)}{\log\,(1 - \sqrt{1-4\alpha^2}) - \log 2\alpha}, & \text{for } R = 3 \text{ and } 0 < \alpha < \alpha_2^* \quad (6.15c) \end{cases}$$

The functions (6.15a) – (6.15c) are shown in Figure 6.8. When we have 1 ring in the network and $\alpha_3^* < \alpha < \widehat{\alpha}$ it follows that nodes with $p(k) < 2$ will always been deemed to be more important than node $N$. In the case of 2 rings and $0 < \alpha < \widehat{\alpha}$ or 3 rings and $0 < \alpha < \alpha_2^*$ we can see that $p(k) < 1$ and therefore node $N$ will always be ranked between nodes 1 and 2 (or equally $m$).

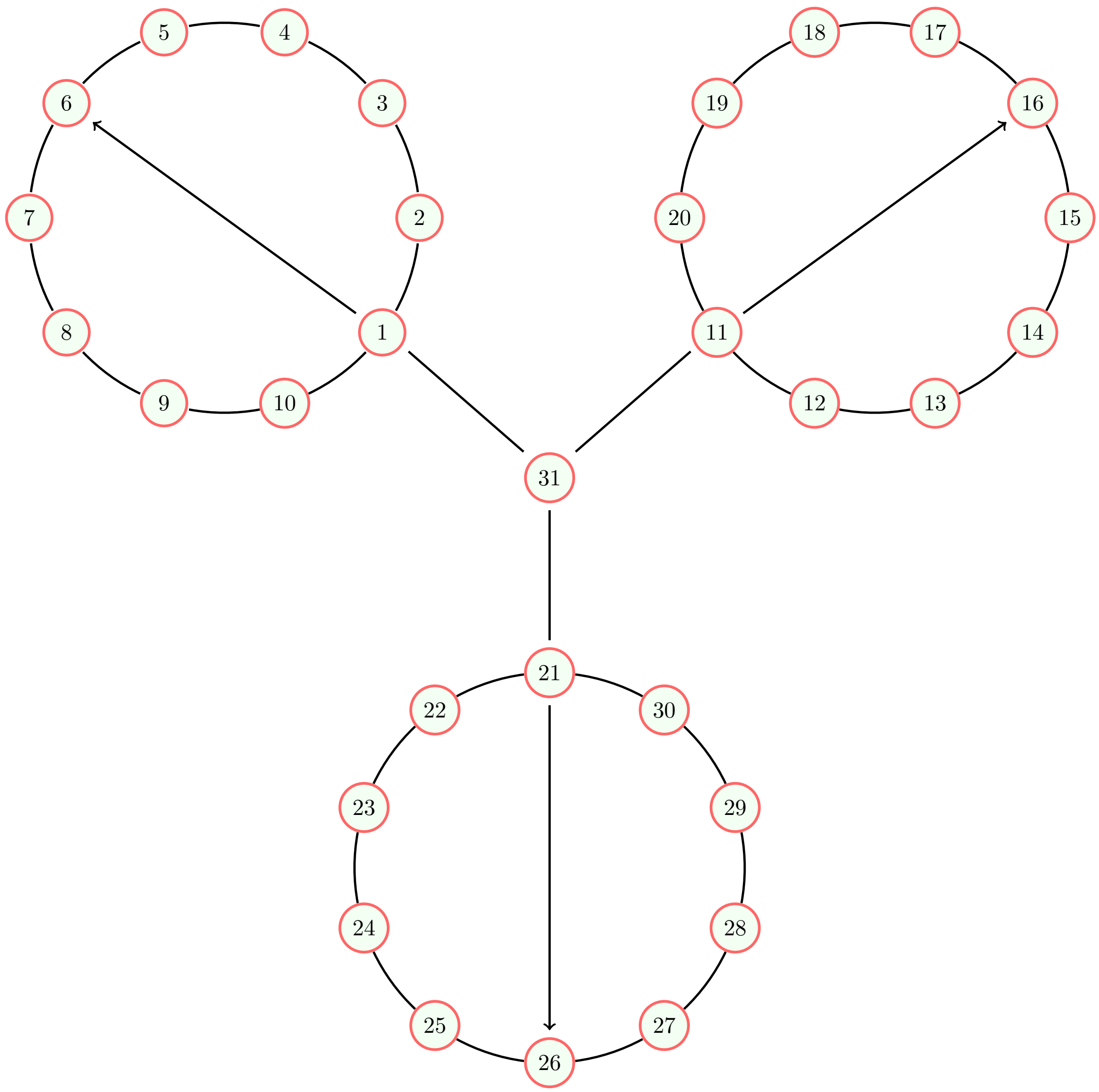
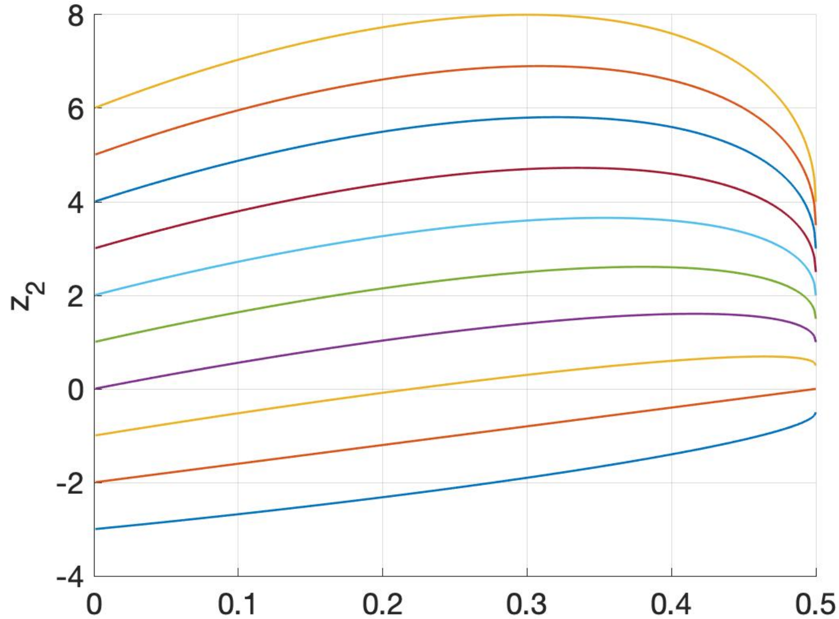
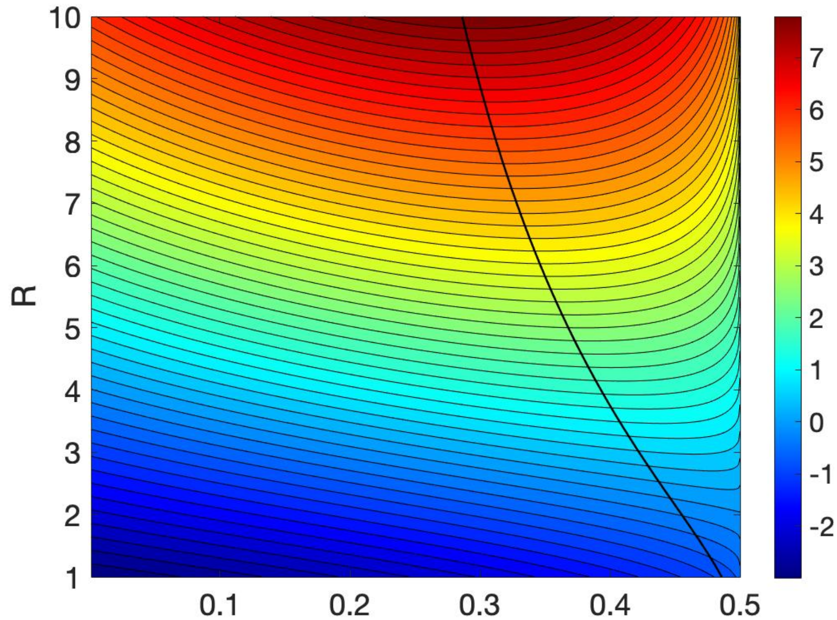



Figure 6.5: *R rings with directed shortcut and undirected hub*: Illustration with directed shortcut and undirected hub ($R = 3$, $m = 10$ and $L = 6$).




$\alpha$

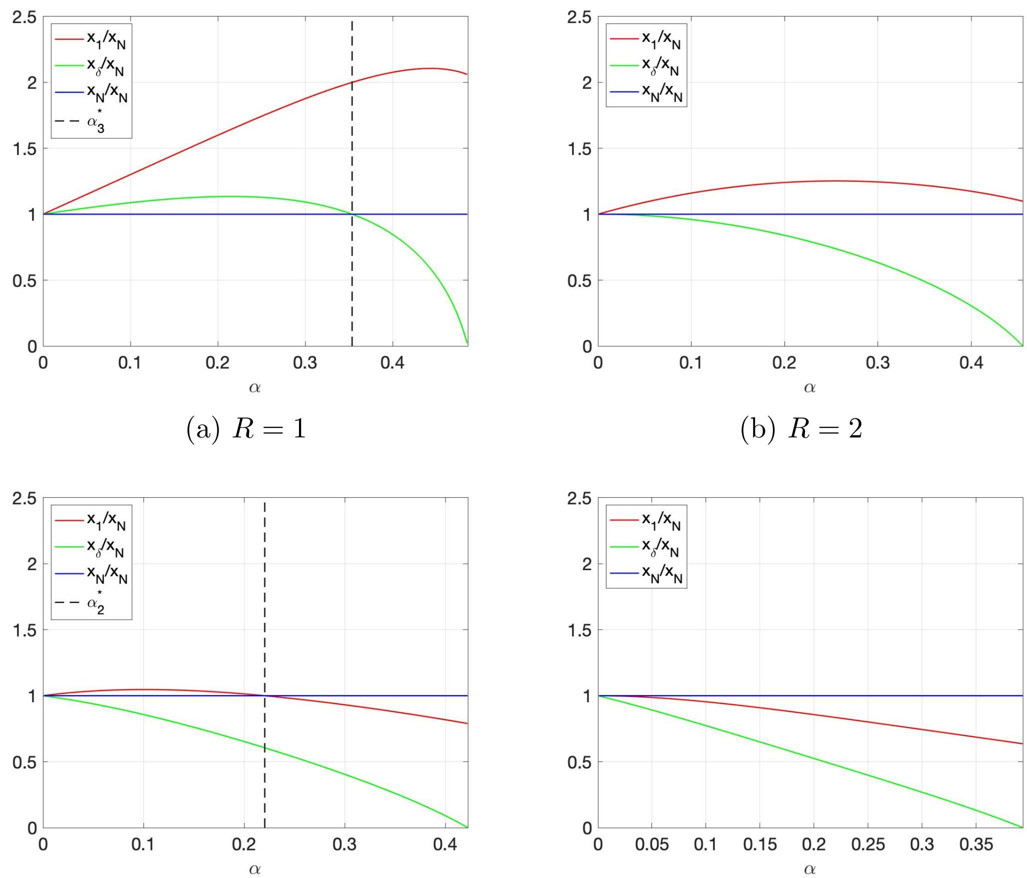
(a) Fixed $R$ ($R$=seq$(1, 10)$). $R = 1$ – bottom curve.




$\alpha$

(b) Contour plot of $z_2$ with the solid black line representing $\alpha = \widehat{\alpha}$.

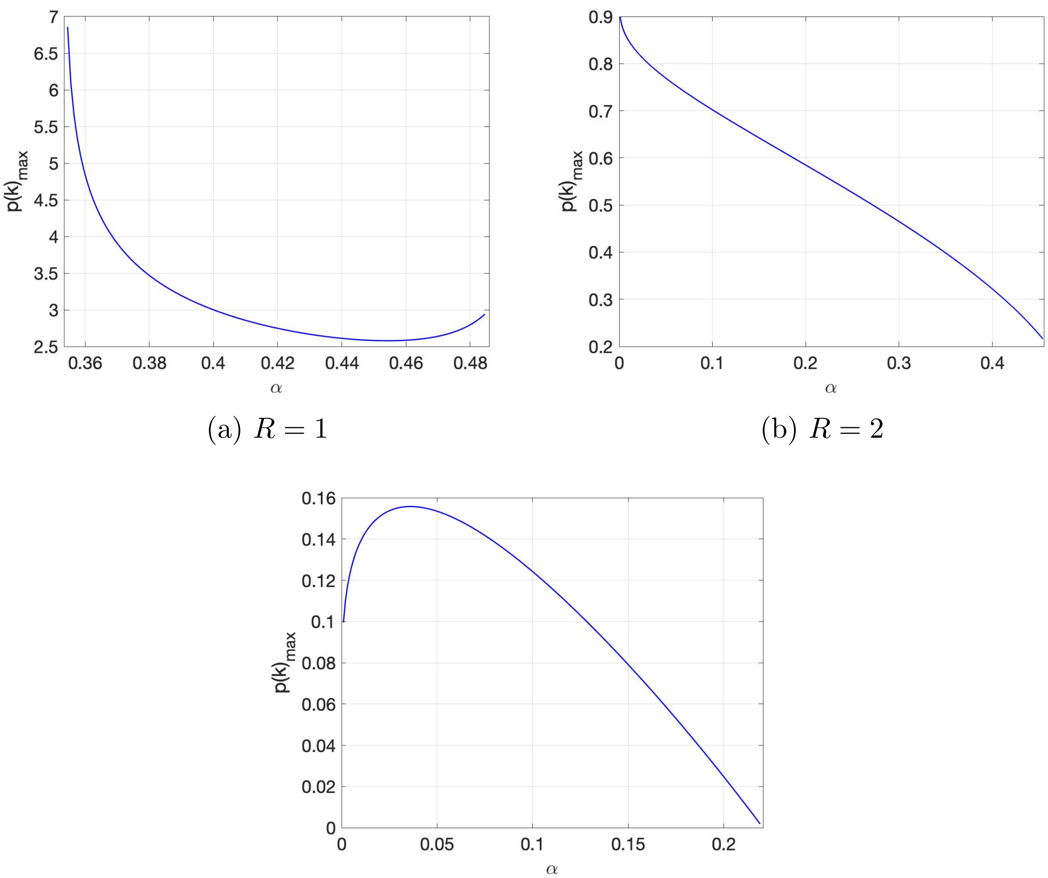
Figure 6.6: *R rings with directed shortcut and undirected hub*: Relationship between number of rings $R$, Katz parameter $\alpha$ and $z_2$ in (6.13)




(a) $R = 1$ (b) $R = 2$

(c) $R = 3$ (d) $R = 4$

Figure 6.7: *R rings with directed shortcut and undirected hub*: Ratio of $x_1$ and $x_\delta$ to $x_N$ for the Katz centrality. We used $m = 1,000$ so $N = 1000R + 1$ with $L = 501$.




(a) $R = 1$

(b) $R = 2$

(c) $R = 3$

Figure 6.8: *R rings with directed shortcut and undirected hub*: $p(k)_{\max}$ indicating the maximum periodic distance for which $x_k > x_N$

## 6.2 $R$-rings with an undirected shortcut

In this section we consider the case of each ring in the network having an undirected shortcut across the ring. As each ring is identical, we denote this using ring 1 as an edge from node 1 to node $L$ and an edge from node $L$ to node 1 where $1 \ll L \leq m/2+1$. This section differs from the previous in that we now consider the ranking of node $L$ due to its additional edge. We will once again, denote the minimum scoring node as node $\delta$.

### 6.2.1 Connection to a directed hub

This subsection considers the case of each of the $R$ rings connecting to the central hub node with an incoming directed edge from central hub node $N$ to each of the local node 1s. Figure 6.9 illustrates an example network with $R = 3$ rings where each ring contains the same identical undirected shortcut from local node 1 to local node 6 (i.e. $1 \mapsto 6, 6 \mapsto 1, 11 \mapsto 16, 16 \mapsto 11, 21 \mapsto 26, 26 \mapsto 21$). Each local node 1 is connected to the central node 31 via an incoming directed edge (i.e. $31 \mapsto 1, 31 \mapsto 11, 31 \mapsto 21$).

For the general $R$-ring network, the Katz system (3.1) reduces to

$$\begin{aligned} x_j - \alpha\left(x_{j-1} + x_{j+1}\right) &= 1, \quad \text{for } 2 \leq j < L \text{ or } L < j \leq m, && (6.16)\\ x_1 - \alpha\left(x_2 + x_m + x_L\right) &= 1, && (6.17)\\ x_L - \alpha\left(x_{L-1} + x_{L+1} + x_1\right) &= 1, && (6.18)\\ x_N - R\alpha x_1 &= 1. && (6.19) \end{aligned}$$

Inserting the ansatz (5.6) we find that, as in the proof of Theorem 4.1, the general equation (6.8) is solved with $b = 1/(1-2\alpha)$ and $t = t_1$. As shown, in the proof of Theorem 5.1, we expect the heights of the spikes at nodes 1 and $L$ to be the same indicating that $h = h'$. We expect to observe geometric decay as we move away from both nodes 1 and nodes $L$. In essence, by considering only the dominant contributions, we have that $x_1 = x_L$ and $x_2 = x_m = x_{L-1} = x_{L+1}$. It therefore follows that (6.17) and (6.18) are equivalent.

Using $x_L \approx b + h$ and solving (6.17) for $h$, we arrive at

$$h \approx \frac{\alpha}{(1-2\alpha)(\sqrt{1-4\alpha^2}-\alpha)}. \tag{6.20}$$

Requiring $h > 0$ in (6.20) places the restriction that $\sqrt{1-4\alpha^2}-\alpha > 0$. Solving for the roots of the quadratic $1 - 5\alpha^2$ we require $0 < \alpha < \widehat{\alpha}$ where

$$\widehat{\alpha} = \frac{1}{\sqrt{5}}. \tag{6.21}$$

We therefore conjecture that the network adjacency matrix has a spectral radius that approaches $\sqrt{5}$ as $N \to \infty$ and hence the Katz system is valid for $0 < \alpha < \widehat{\alpha}$. Table 6.5 gives an indication of the discrepancy between the spectral radius of the adjacency matrix ($\rho(A)$) and the reciprocal of the upper bound placed on the $\alpha$ domain ($1/\widehat{\alpha}$) such that $h > 0$.

| $R$ | $\widehat{\alpha}$ | $1/\widehat{\alpha}$ | $\lvert\rho(A) - 1/\widehat{\alpha}\rvert$ | | |
|---|---|---|---|---|---|
| | | | $m = 10^2$ | $m = 10^4$ | $m = 10^6$ |
| 1 | $1/\sqrt{5}$ | $\sqrt{5}$ | $\approx 3.2 \times 10^{-11}$ | $\approx 1.9 \times 10^{-14}$ | $\approx 1.3 \times 10^{-14}$ |
| 2 | $1/\sqrt{5}$ | $\sqrt{5}$ | $\approx 3.2 \times 10^{-11}$ | $\approx 1.5 \times 10^{-14}$ | $\approx 9.8 \times 10^{-15}$ |
| 3 | $1/\sqrt{5}$ | $\sqrt{5}$ | $\approx 3.2 \times 10^{-11}$ | $\approx 1.9 \times 10^{-14}$ | $\approx 2.0 \times 10^{-14}$ |
| 4 | $1/\sqrt{5}$ | $\sqrt{5}$ | $\approx 3.2 \times 10^{-11}$ | $\approx 7.5 \times 10^{-15}$ | $\approx 1.6 \times 10^{-14}$ |

Table 6.5: *$R$ rings with undirected shortcut and directed hub*: Comparison of $\rho(A)$ and $1/\widehat{\alpha}$ where $\widehat{\alpha}$ is the upper bound of the domain defined by the heuristic analysis. [$\rho(A)$ was computed using MATLAB's `eigs`.]

We know that $x_1 = b + h$ and from (6.19) that $x_N = 1 + R\alpha(b+h)$. We now turn our attention to when $x_N > x_1$, i.e.

$$1 + (R\alpha - 1)(b + h) > 0.$$

Using the expression for $h$ in (6.20) this gives

$$1 + (R\alpha - 1)\left[\frac{1}{1-2\alpha} + \frac{\alpha}{(1-2\alpha)(\sqrt{1-4\alpha^2}-\alpha)}\right] > 0$$

which rearranges to

$$(1-2\alpha)(\sqrt{1-4\alpha^2}-\alpha)+(R\alpha-1)\sqrt{1-4\alpha^2}>0.$$

We arrive at

$$z_3 := (R-2)\sqrt{1-4\alpha^2}+2\alpha-1>0. \tag{6.22}$$

From Figure 6.10 we observe that $z_3 < 0$ for $R \leq 2$ and $z_2 > 0$ for $R \geq 3$ over the domain for which the Katz system is valid. We can therefore conclude that $x_N > x_1$ when we have 3 or more rings in our network and $x_1 > x_N$ when we have fewer than 3 rings.

When we have fewer than 3 rings in the network we have node 1 being ranked as more important than node $N$. A natural extension is to evaluate how $x_N$ compares with other nodes in the ring. To do so we once again, first compare with node $\delta$. Using $x_\delta \approx b$ we find $x_N > x_\delta$ when

$$1+R\alpha(b+h)>b.$$

Using the expression for $h$ in (6.20) this gives

$$1+R\alpha\left[\frac{1}{1-2\alpha}+\frac{\alpha}{(1-2\alpha)(\sqrt{1-4\alpha^2}-\alpha)}\right]>\frac{1}{1-2\alpha}$$

which rearranges to

$$(1-2\alpha)(\sqrt{1-4\alpha^2}-\alpha)+R\alpha\sqrt{1-4\alpha^2}>\sqrt{1-4\alpha^2}-\alpha.$$

We arrive at

$$2\alpha+(R-2)\sqrt{1-4\alpha^2}>0. \tag{6.23}$$

It is clear that in the case when $R=2$, (6.23) is satisfied for $0<\alpha<\widehat{\alpha}$ indicating that $x_N > x_\delta$. In the case when $R=1$, we require $2\alpha-\sqrt{1-4\alpha^2}>0$ so we identify a pole $\alpha_4^*=1/\sqrt{8}$ where (6.23) changes sign. We can therefore only satisfy (6.23) for $\alpha_4^*<\alpha<\widehat{\alpha}$, meaning that $x_\delta > x_N$ when $0<\alpha<\alpha_4^*$.

A summary of findings can be found below in Table 6.6.

| R | Degree | Katz | Eigenvector |
|---|---|---|---|
| 1 | $x_1 > x_\delta > x_N$ | $x_1 > x_\delta > x_N \ \forall \alpha \in (0, \alpha_4^*)$<br>$x_1 > x_N > x_\delta \ \forall \alpha \in (\alpha_4^*, \widehat{\alpha})$ | $x_1 > x_N > x_\delta$ |
| 2 | $x_1 > x_N = x_\delta$ | $x_1 > x_N > x_\delta \ \forall \alpha \in (0, \widehat{\alpha})$ | $x_1 > x_N > x_\delta$ |
| 3 | $x_N = x_1 > x_\delta$ | $x_N > x_1 > x_\delta \ \forall \alpha \in (0, \widehat{\alpha})$ | $x_N > x_1 > x_\delta$ |
| $\geq 4$ | $x_N > x_1 > x_\delta$ | $x_N > x_1 > x_\delta \ \forall \alpha \in (0, \widehat{\alpha})$ | $x_N > x_1 > x_\delta$ |

Table 6.6: *R rings with undirected shortcut and directed hub*: Katz centrality for $x_1, x_\delta$ and $x_N$ from the heuristic analysis compared with degree ($\alpha \to 0$) and eigenvector ($\alpha \to 1/\rho(A)$) centrality

Figure 6.11 shows the ratio of nodes 1 and $\delta$ to node $N$ as $\alpha$ is varied for $1 \leq R \leq 4$. To do this, we used $m = 1,000$ with $L = 501$ and solved the Katz system (3.1) directly. By examining these figures we can see that the results from our heuristic analysis, presented in Table 6.6, agree with the solution of the Katz system. In particular Figure 6.11a shows the change in ranking of nodes $\delta$ and 1 on passing the threshold $\alpha_4^*$.

In the case when node $N$ is neither ranked first nor last, i.e. when $x_1 > x_N > x_\delta$, we can quantify where node $N$ is ranked in relation to nodes in the ring by considering where $x_k > x_N$. This condition can be reduced to

$$b + h t_1^{p_{1,L}(k)} > 1 + R\alpha(b + h).$$

Using the expression for $h$ in (6.20) this gives

$$\frac{\alpha}{(1-2\alpha)(\sqrt{1-4\alpha^2}-\alpha)} \left[ t_1^{p_{1,L}(k)} - R\alpha \right] > 1 + \frac{R\alpha - 1}{1 - 2\alpha}$$

which rearranges to

$$t_1^{p_{1,L}(k)} - R\alpha > (R-2)(\sqrt{1-4\alpha^2} - \alpha).$$

We arrive at

$$t_1^{p_{1,L}(k)} > 2\alpha + (R-2)\sqrt{1-4\alpha^2},$$

where we find that $p_{1,L}(k) < p_{1,L}(k)_{\max}$ for

$$p_{1,L}(k)_{\max} = \begin{cases} \dfrac{\log(2\alpha - \sqrt{1-4\alpha^2})}{\log(1-\sqrt{1-4\alpha^2}) - \log 2\alpha}, & \text{for } R = 1 \text{ and } \alpha_4^* < \alpha < \widehat{\alpha} \quad (6.24a) \\ \dfrac{\log(2\alpha)}{\log(1-\sqrt{1-4\alpha^2}) - \log 2\alpha}, & \text{for } R = 2 \text{ and } 0 < \alpha < \widehat{\alpha} \quad (6.24b) \end{cases}$$

The functions (6.24a) and (6.24b) are shown in Figure 6.12. In the case of 1 ring in the network, when $\alpha_4^* < \alpha < \widehat{\alpha}$, it follows that nodes with $p_{1,L}(k) \leq 1$, that is nodes $2, m, L-1$ and $L+1$ will always be considered more important that node $N$. In the case of 2 rings, since $p_{1,L}(k) < 1$ it follows that node $N$ will be ranked between nodes 1 and 2 (or equally $m$, $L-1$ or $L+1$).

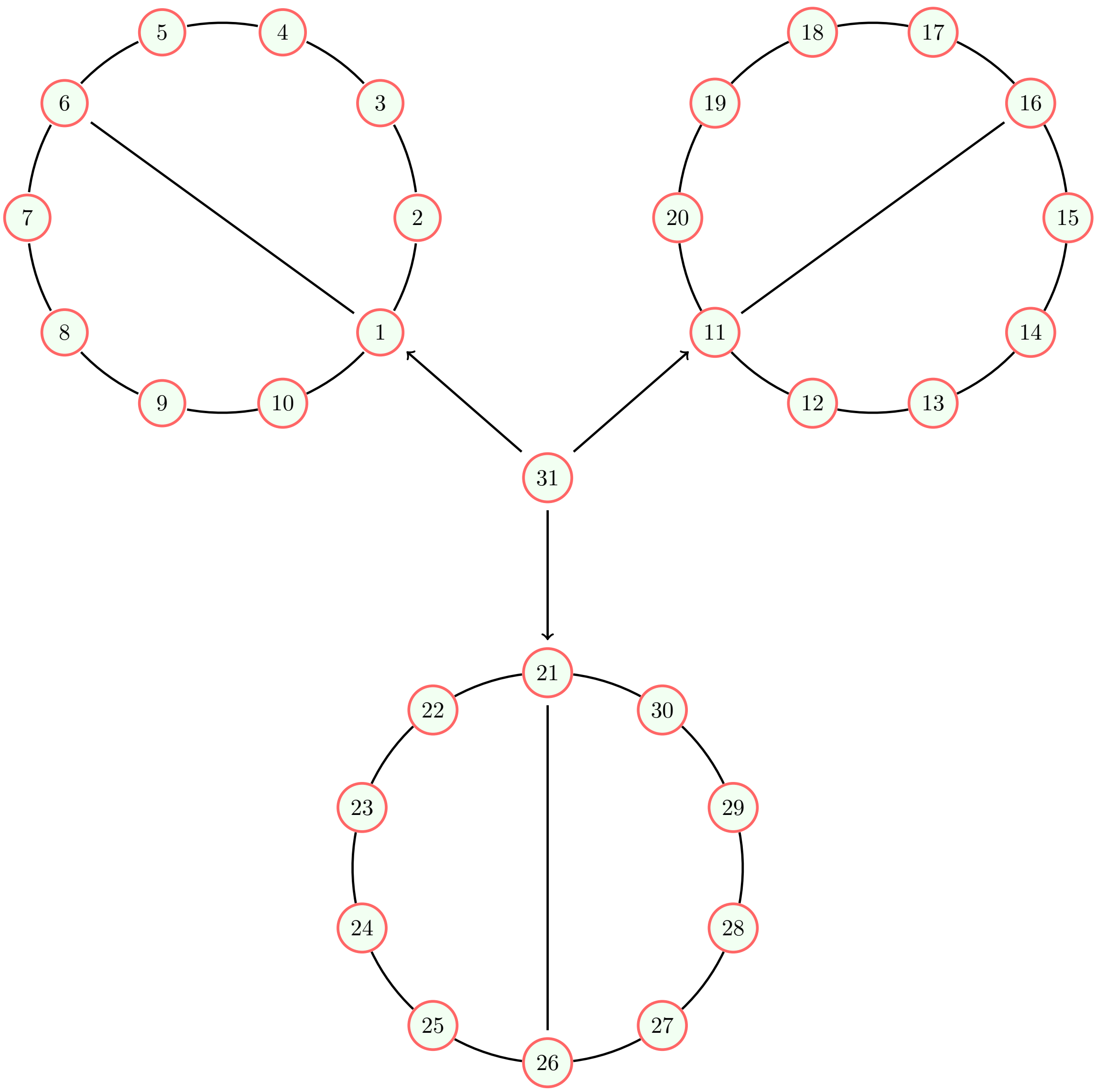
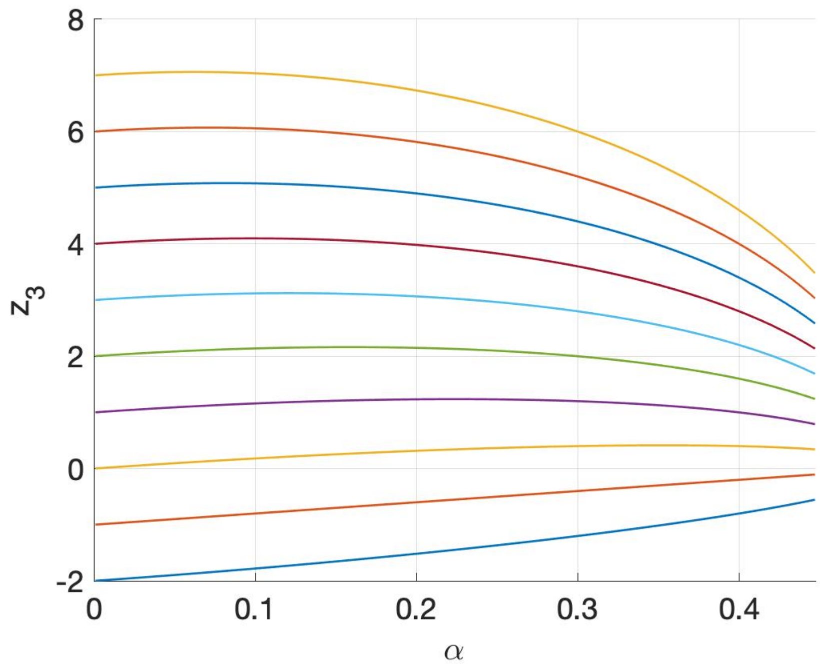
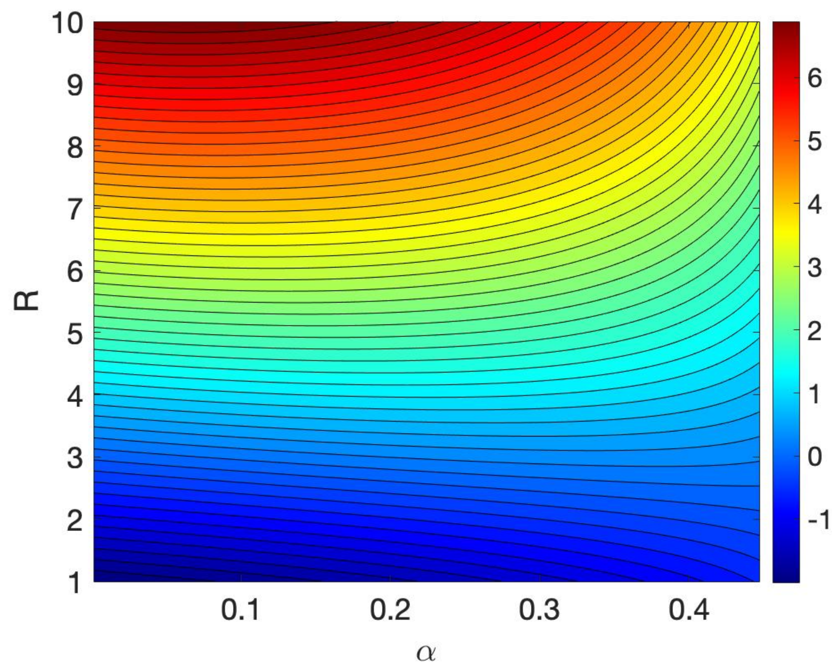



Figure 6.9: *R rings with undirected shortcut and directed hub*: Illustration with undirected shortcut and directed hub ($R = 3$, $m = 10$ and $L = 6$).

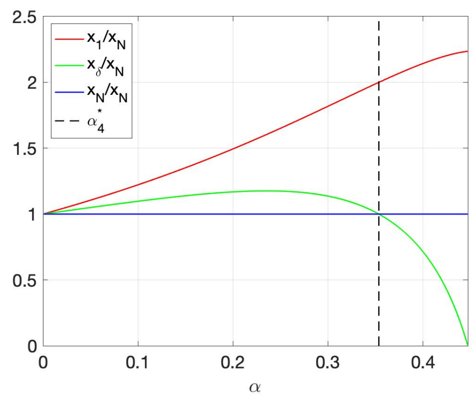



(a) Fixed $R$ ($R$=seq$(1, 10)$). $R = 1$ – bottom curve.

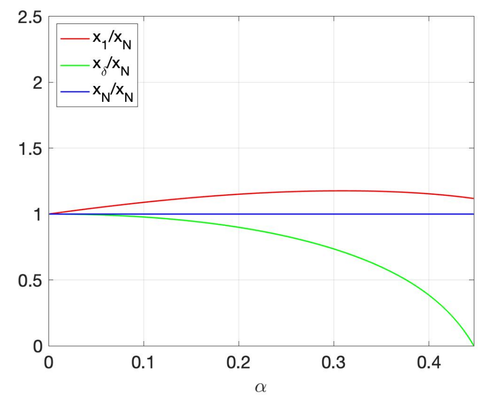



(b) Contour plot of $z_3$.

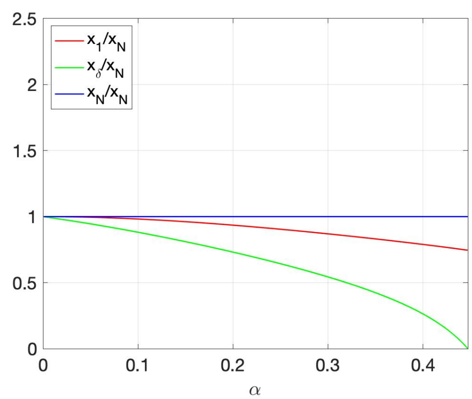
Figure 6.10: *R rings with undirected shortcut and directed hub*: Relationship between number of rings $R$, Katz parameter $\alpha$ and $z_3$ in (6.22)



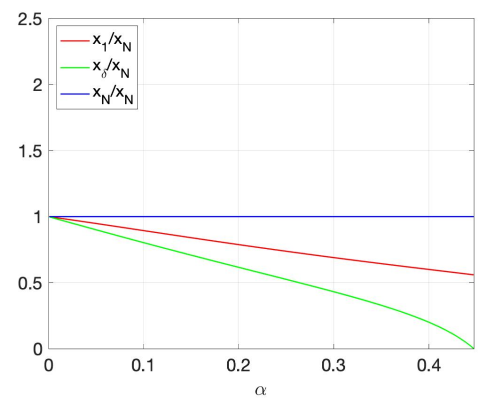

(a) $R = 1$




(b) $R = 2$

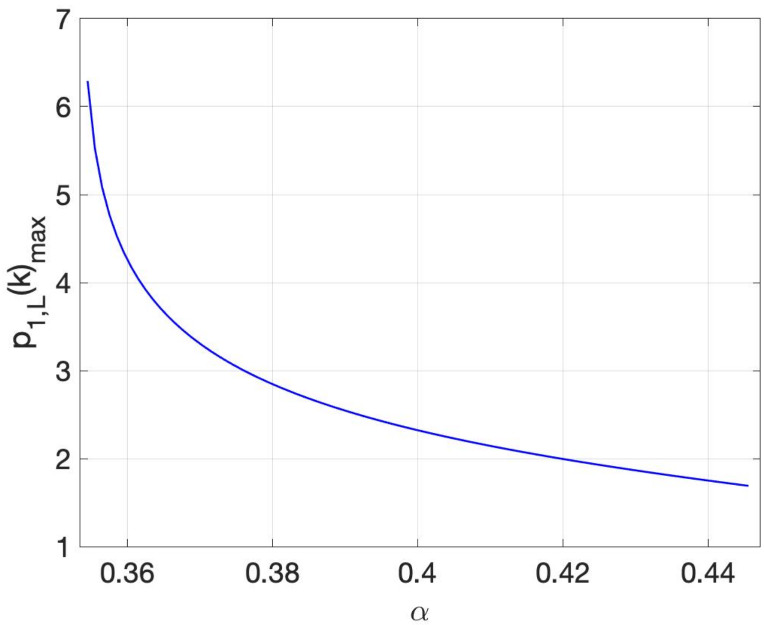



(c) $R = 3$

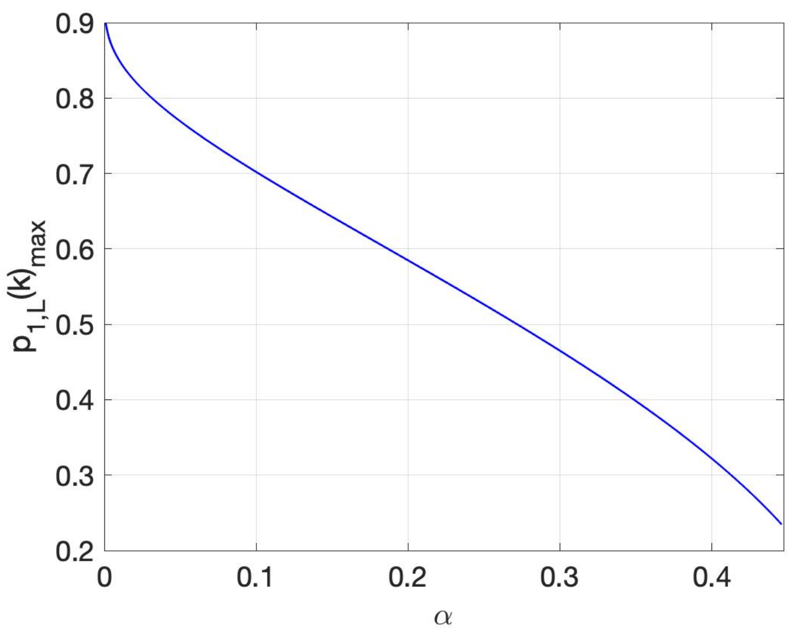



(d) $R = 4$

Figure 6.11: *$R$ rings with undirected shortcut and directed hub*: Ratio of $x_1$ and $x_\delta$ to $x_N$ for the Katz centrality. We used $m = 1,000$ so $N = 1000R + 1$ with $L = 501$.




(a) $R = 1$




(b) $R = 2$

Figure 6.12: *R rings with undirected shortcut and directed hub*: $p_{1,L}(k)_{\max}$ indicating the maximum periodic distance for which $x_k > x_N$

### 6.2.2 Connection to an undirected hub

This subsection considers the case of each of the $R$ rings connecting to the central hub node with an undirected edge from central hub node $N$ to each of the local node 1s. Figure 6.13 illustrates an example network with $R = 3$ rings where each ring contains the same identical undirected shortcut from local node 1 to local node 6 (i.e. $1 \mapsto 6, 6 \mapsto 1, 11 \mapsto 16, 16 \mapsto 11, 21 \mapsto 26, 26 \mapsto 21$). Each local node 1 is connected to the central node 31 via an undirected edge (i.e. $1 \mapsto 31, 31 \mapsto 1, 11 \mapsto 31, 31 \mapsto 11, 21 \mapsto 31, 31 \mapsto 21$).

For the general $R$-ring network, the Katz system (3.1) reduces to

$$\begin{aligned} x_j - \alpha\,(x_{j-1} + x_{j+1}) &= 1, \qquad \text{for } 2 \leq j < L \text{ or } L < j \leq m, && (6.25)\\ x_1 - \alpha\,(x_2 + x_m + x_L + x_N) &= 1, && (6.26)\\ x_L - \alpha\,(x_{L-1} + x_{L+1} + x_1) &= 1, && (6.27)\\ x_N - R\alpha x_1 &= 1. && (6.28) \end{aligned}$$

Inserting the ansatz (5.6) we find that, as in the proof of Theorem 4.1, the general equation (6.25) is solved with $b = 1/(1 - 2\alpha)$ and $t = t_1$. In this instance, we would not expect the height of the spikes at nodes 1 and $L$ to be the same due to the influence of the additional outgoing edge at node 1 towards node $N$. However, we will continue to observe the same geometric decay rates around nodes 1 and $L$.

Using (6.26), (6.27) and (6.28) we arrive at the system

$$\begin{aligned} (1-2\alpha)(\sqrt{1-4\alpha^2} - R\alpha^2)h - \alpha(1-2\alpha)h' &= \alpha\,[2 + (R-2)\alpha] && (6.29)\\ -\alpha(1-2\alpha)h + (1-2\alpha)\sqrt{1-4\alpha^2}h' &= \alpha && (6.30) \end{aligned}$$

for $h$ and $h'$. We now use (6.30) to express $h'$ in terms of $h$ so that

$$h' = \frac{\alpha\,[1 + (1-2\alpha)h]}{(1-2\alpha)\sqrt{1-4\alpha^2}}$$

and use this in (6.29). Following some algebraic manipulation we arrive at

$$h \approx \frac{\alpha\left[(2+(R-2)\alpha)\sqrt{1-4\alpha^2}+\alpha\right]}{(1-2\alpha)(1-5\alpha^2-R\alpha^2\sqrt{1-4\alpha^2})} \tag{6.31}$$

$$h' \approx \frac{\alpha\left[2\alpha(1-\alpha)+\sqrt{1-4\alpha^2}\right]}{(1-2\alpha)(1-5\alpha^2-R\alpha^2\sqrt{1-4\alpha^2})} \tag{6.32}$$

We know that the upper bound on the $\alpha$ domain must be less than 1/2 and therefore requiring $h > 0$ and $h' > 0$ means we need the denominators of (6.31) and (6.32) to be positive. This places the restriction

$$1-5\alpha^2-R\alpha^2\sqrt{1-4\alpha^2} > 0 \tag{6.33}$$

which when solved will constrict $0 < \alpha < \widehat{\alpha}$. We conjecture that the network adjacency matrix has a spectral radius that approaches $1/\widehat{\alpha}$ as $N \to \infty$. Table 6.7 gives an indication of the discrepancy between the spectral radius of the adjacency matrix ($\rho(A)$) and the reciprocal of the upper bound placed on the $\alpha$ domain ($1/\widehat{\alpha}$) such that $h > 0$.

| $R$ | $\widehat{\alpha}$ | $1/\widehat{\alpha}$ | $\lvert\rho(A)-1/\widehat{\alpha}\rvert$ | | |
|---|---|---|---|---|---|
| | | | $m = 10^2$ | $m = 10^4$ | $m = 10^6$ |
| 1 | $\approx 0.4254$ | $\approx 2.3506$ | $\approx 2.5 \times 10^{-13}$ | $\approx 3.6 \times 10^{-15}$ | $\approx 0$ |
| 2 | $\approx 0.4019$ | $\approx 2.4879$ | $\approx 1.8 \times 10^{-15}$ | $\approx 5.8 \times 10^{-15}$ | $\approx 2.7 \times 10^{-15}$ |
| 3 | $\approx 0.3792$ | $\approx 2.6373$ | $\approx 1.8 \times 10^{-15}$ | $\approx 6.2 \times 10^{-15}$ | $\approx 2.7 \times 10^{-15}$ |
| 4 | $\approx 0.3583$ | $\approx 2.7911$ | $\approx 8.9 \times 10^{-16}$ | $\approx 1.8 \times 10^{-15}$ | $\approx 3.1 \times 10^{-15}$ |

Table 6.7: *R rings with undirected shortcut and undirected hub*: Comparison of $\rho(A)$ and $1/\widehat{\alpha}$ where $\widehat{\alpha}$ is the upper bound of the domain defined by the heuristic analysis. [$\rho(A)$ was computed using MATLAB's `eigs`.]

We know that $x_1 = b+h$ and from (6.28) that $x_N = 1+R\alpha(b+h)$. We now turn our attention to when $x_N > x_1$, i.e.

$$1+(R\alpha-1)(b+h) > 0.$$

Using the expression for $h$ in (6.31) this gives

$$1 + (R\alpha - 1)\left[\frac{1}{1-2\alpha} + \frac{\alpha\left[(2+(R-2)\alpha)\sqrt{1-4\alpha^2}+\alpha\right]}{(1-2\alpha)(1-5\alpha^2-R\alpha^2\sqrt{1-4\alpha^2})}\right] > 0$$

which rearranges to

$$(R-2)(1-5\alpha^2-R\alpha^2\sqrt{1-4\alpha^2})+(R\alpha-1)(2+(R-2)\alpha)\sqrt{1-4\alpha^2}+\alpha(R\alpha-1) > 0.$$

We arrive at

$$z_4 := \left[(R+2)\alpha - 2\right]\sqrt{1-4\alpha^2} + 2\alpha^2(5-2R) - \alpha + R - 2 > 0. \tag{6.34}$$

From Figure 6.14 we observe that $z_4 < 0$ for $R \leq 2$ and $z_4 > 0$ for $R \geq 4$ over the domain for which the Katz system is valid. We can therefore conclude that $x_N > x_1$ when we have 4 or more rings in our network and $x_1 > x_N$ when we have 2 or fewer rings.

In the case of having $R = 3$ rings, we require $(5\alpha-2)\sqrt{1-4\alpha^2} - 2\alpha^2 - \alpha + 1 > 0$ and so solving for the roots we observe that $z_4$ changes sign on crossing a pole

$$\begin{aligned}\alpha_5^* &= \frac{\sqrt{14}\sqrt{3}}{13}\sin\left(-\frac{1}{3}\arctan\left(\frac{13\sqrt{7}}{49}\right)+\frac{\pi}{3}\right)\\ &\quad -\frac{\sqrt{14}}{13}\sin\left(\frac{1}{3}\arctan\left(\frac{13\sqrt{7}}{49}\right)+\frac{\pi}{6}\right)+\frac{1}{13} \approx 0.2578.\end{aligned}$$

We conjecture that this is the threshold value beyond which node $N$ is regarded as more central than node 1.

Given that node 1 has an edge more than node $L$ we expect that $x_1 > x_L$ and therefore $h > h'$. We now therefore consider if $x_N > x_L$ in the case where $x_1 > x_N$, that is, when $R = 1$, $R = 2$ or $R = 3$ and $0 < \alpha < \alpha_5^*$.

Considering the condition $x_N > x_L$ we have

$$1 + R\alpha(b+h) > b + h'.$$

Using the expression for $h$ in (6.31) and $h'$ in (6.32) this gives

$$1 + R\alpha\left[\frac{1}{1-2\alpha} + \frac{\alpha\left[(2+(R-2)\alpha)\sqrt{1-4\alpha^2}+\alpha\right]}{(1-2\alpha)(1-5\alpha^2-R\alpha^2\sqrt{1-4\alpha^2})}\right] - \left[\frac{1}{1-2\alpha} + \frac{\alpha\left[2\alpha(1-\alpha)+\sqrt{1-4\alpha^2}\right]}{(1-2\alpha)(1-5\alpha^2-R\alpha^2\sqrt{1-4\alpha^2})}\right] > 0$$

which rearranges to

$$(R-2)(1-5\alpha^2-R\alpha^2\sqrt{1-4\alpha^2}) + R\alpha\left[(2+(R-2)\alpha)\sqrt{1-4\alpha^2}+\alpha\right] - 2\alpha(1-\alpha) - \sqrt{1-4\alpha^2} > 0.$$

We arrive at

$$z_5 := (2R\alpha-1)\sqrt{1-4\alpha^2} + 4(3-R)\alpha^2 - 2\alpha + R - 2 > 0. \tag{6.35}$$

From Figure 6.15 we observe that $z_5 > 0$ for $R \geq 3$ and $z_5 < 0$ for $R = 1$ over $0 < \alpha < \widehat{\alpha}$ and therefore it follows that $x_N > x_L$ when we have 3 or more rings and $x_L > x_N$ when we have one ring. However, in the case when $R = 2$ we observe a critical choice for $\alpha$ where $z_5$ changes sign.

When $R = 2$, from (6.35) we have the condition $(4\alpha-1)\sqrt{1-4\alpha^2}-2\alpha(1-2\alpha) > 0$. Solving for the roots we observe that $z_5$ changes sign on crossing the pole

$$\alpha_6^* = \frac{\sqrt{46}}{15}\sin\left(\frac{1}{3}\arctan\left(\frac{15\sqrt{111}}{269}\right) + \frac{\pi}{6}\right) + \frac{1}{30} \approx 0.3248.$$

We conjecture that this is the threshold value beyond which node $N$ is regarded as more central than node $L$.

In the case when node $N$ is not regarded as more central than node 1, a natural extension is to see how $x_N$ compares with the other nodes in the ring. To do so we first compare with node $\delta$ (the minimum scored node). Using $x_\delta \approx b$ we find $x_N > x_\delta$ when

$$1 + R\alpha(b+h) > b.$$

Using the expression for $h$ in (6.31) this gives

$$1 + R\alpha \left[ \frac{1}{1-2\alpha} + \frac{\alpha\left[(2+(R-2)\alpha)\sqrt{1-4\alpha^2}+\alpha\right]}{(1-2\alpha)(1-5\alpha^2-R\alpha^2\sqrt{1-4\alpha^2})} \right] > \frac{1}{1-2\alpha}$$

which rearranges to

$$(R-2)(1-5\alpha^2-R\alpha^2\sqrt{1-4\alpha^2}) + R\alpha\left[(2+(R-2)\alpha)\sqrt{1-4\alpha^2}+\alpha\right] > 0.$$

We arrive at

$$z_6 := 2R\alpha\sqrt{1-4\alpha^2} + 2(5-2R)\alpha^2 + R - 2 > 0. \tag{6.36}$$

From Figure 6.16 we observe that $z_6 > 0$ for $R \geq 2$ and that when $R = 1$ there is a pole where $z_6$ changes sign. In the case of $R = 1$ we solve for the roots of $2\alpha\sqrt{1-4\alpha^2} + 6\alpha^2 - 1$ which gives

$$\alpha_7^* \approx 0.2953.$$

We claim this is the threshold value beyond which node $N$ is regarded as more important than the minimum scored node in the ring (node $\delta$).

A summary of findings can be found below in Table 6.8.

| R | Degree | Katz | Eigenvector |
|---|---|---|---|
| 1 | $x_1 > x_L > x_\delta > x_N$ | $x_1 > x_L > x_\delta > x_N \ \forall\alpha \in (0, \alpha_7^*)$ | $x_1 > x_L > x_N > x_\delta$ |
| | | $x_1 > x_L > x_N > x_\delta \ \forall\alpha \in (\alpha_7^*, \widehat{\alpha})$ | |
| 2 | $x_1 > x_L > x_N = x_\delta$ | $x_1 > x_L > x_N > x_\delta \ \forall\alpha \in (0, \alpha_6^*)$ | $x_1 > x_N > x_L > x_\delta$ |
| | | $x_1 > x_N > x_L > x_\delta \ \forall\alpha \in (\alpha_6^*, \widehat{\alpha})$ | |
| 3 | $x_1 > x_N = x_L > x_\delta$ | $x_1 > x_N > x_L > x_\delta \ \forall\alpha \in (0, \alpha_5^*)$ | $x_N > x_1 > x_L > x_\delta$ |
| | | $x_N > x_1 > x_L > x_\delta \ \forall\alpha \in (\alpha_5^*, \widehat{\alpha})$ | |
| $\geq 4$ | $x_N \geq x_1 > x_L > x_\delta$ | $x_N > x_1 > x_L > x_\delta \ \forall\alpha \in (0, \widehat{\alpha})$ | $x_N > x_1 > x_L > x_\delta$ |

Table 6.8: *R rings with undirected shortcut and undirected hub*: Katz centrality for $x_1, x_L, x_\delta$ and $x_N$ from the heuristic analysis compared with degree ($\alpha \to 0$) and eigenvector ($\alpha \to 1/\rho(A)$) centrality

Figure 6.17 shows the ratio of nodes 1, $L$ and $\delta$ to node $N$ as $\alpha$ is varied for

$1 \leq R \leq 4$. To do this, we used $m = 1,000$ with $L = 501$ and solved the Katz system (3.1) directly. By examining these figures we can see that the results from our heuristic analysis, presented in Table 6.8, agree with the solution of the Katz system. In particular:

- Figure 6.17a shows the change in ranking of nodes $\delta$ and $N$ on passing the threshold $\alpha_7^*$,
- Figure 6.17b shows the change in ranking of nodes $L$ and $N$ on passing the threshold $\alpha_6^*$,
- Figure 6.17c shows the change in ranking of nodes 1 and $N$ on passing the threshold $\alpha_5^*$.

Now when $x_1 > x_N > x_\delta$ and $x_L > x_N > x_\delta$ we may quantify where node $N$ is ranked by considering where $x_k > x_N$. We fist examine this inequality for nodes close to node 1, where we have

$$b + ht_1^{p(k)} > 1 + R\alpha(b + h).$$

Using the expression for $h$ in (6.31) this gives

$$\frac{\alpha\left[(2+(R-2)\alpha)\sqrt{1-4\alpha^2}+\alpha\right]}{(1-2\alpha)(1-5\alpha^2-R\alpha^2\sqrt{1-4\alpha^2})}\left[t_1^{p(k)}-R\alpha\right] > 1 + \frac{R\alpha-1}{1-2\alpha}$$

which rearranges to

$$\left[(2+(R-2)\alpha)\sqrt{1-4\alpha^2}+\alpha\right]\left[t_1^{p(k)}-R\alpha\right] > (R-2)(1-5\alpha^2-R\alpha^2\sqrt{1-4\alpha^2}).$$

We arrive at

$$t_1^{p(k)} > \frac{2R\alpha\sqrt{1-4\alpha^2}+2(5-2R)\alpha^2+R-2}{(2+(R-2)\alpha)\sqrt{1-4\alpha^2}+\alpha},$$

where we find that $p(k) < p(k)_{\max}$ for

$$p(k)_{\max} = \begin{cases} \dfrac{\log\left(2\alpha\sqrt{1-4\alpha^2}+6\alpha^2-1\right)-\log\left((2-\alpha)\sqrt{1-4\alpha^2}+\alpha\right)}{\log\left(1-\sqrt{1-4\alpha^2}\right)-\log 2\alpha}, & \text{(6.37a)} \\ \dfrac{\log\left(4\alpha\sqrt{1-4\alpha^2}+2\alpha^2\right)-\log\left(2\sqrt{1-4\alpha^2}+\alpha\right)}{\log\left(1-\sqrt{1-4\alpha^2}\right)-\log 2\alpha}, & \text{(6.37b)} \\ \dfrac{\log\left(6\alpha\sqrt{1-4\alpha^2}-2\alpha^2+1\right)-\log\left((2+\alpha)\sqrt{1-4\alpha^2}+\alpha\right)}{\log\left(1-\sqrt{1-4\alpha^2}\right)-\log 2\alpha}, & \text{(6.37c)} \end{cases}$$

where (6.37a) is valid for $R = 1$ and $\alpha_7^* < \alpha < \widehat{\alpha}$, (6.37b) is valid for $R = 2$ and (6.37c) is valid for $R = 3$ and $0 < \alpha < \alpha_5^*$.

The functions (6.37a) – (6.37c) are shown in Figure 6.18. In the case of $R = 1$, assuming we do not choose $\alpha$ really close to $\alpha_7^*$ then we find that node $N$ is, at worse, ranked 6 positions less than node 1. In the case when $R = 2$ or $R = 3$ we see that $p(k)_{\max} < 1$ indicating that node $N$ will be ranked more important than node 2 (or equivalently node $m$).

We now examine this inequality, $x_k > x_N$, for nodes close to node $L$, where we have

$$b + h' t_1^{p_L(k)} > 1 + R\alpha(b+h).$$

Using the expression for $h$ in (6.31) and $h'$ in (6.32) this gives

$$\frac{\alpha\left[2\alpha(1-\alpha)+\sqrt{1-4\alpha^2}\right]}{(1-2\alpha)(1-5\alpha^2-R\alpha^2\sqrt{1-4\alpha^2})}\left[t_1^{p_L(k)}\right]$$
$$> 1 + \frac{R\alpha-1}{1-2\alpha} + \frac{R\alpha^2\left[(2+(R-2)\alpha)\sqrt{1-4\alpha^2}+\alpha\right]}{(1-2\alpha)(1-5\alpha^2-R\alpha^2\sqrt{1-4\alpha^2})}$$

which rearranges to

$$\left[2\alpha(1-\alpha)+\sqrt{1-4\alpha^2}\right]\left[t_1^{p_L(k)}\right]$$
$$> (R-2)(1-5\alpha^2-R\alpha^2\sqrt{1-4\alpha^2}) + R\alpha\left[(2+(R-2)\alpha)\sqrt{1-4\alpha^2}+\alpha\right].$$

We arrive at

$$t_1^{p_L(k)} > \frac{2R\alpha\sqrt{1-4\alpha^2} + 2(5-2R)\alpha^2 + R - 2}{2\alpha(1-\alpha) + \sqrt{1-4\alpha^2}},$$

where we find that $p_L(k) < p_L(k)_{\max}$ for

$$p_L(k)_{\max} = \begin{cases} \dfrac{\log\left(2\alpha\sqrt{1-4\alpha^2} + 6\alpha^2 - 1\right) - \log\left(2\alpha(1-\alpha) + \sqrt{1-4\alpha^2}\right)}{\log\left(1 - \sqrt{1-4\alpha^2}\right) - \log 2\alpha}, & (6.38a) \\ \dfrac{\log\left(4\alpha\sqrt{1-4\alpha^2} + 2\alpha^2\right) - \log\left(2\alpha(1-\alpha) + \sqrt{1-4\alpha^2}\right)}{\log\left(1 - \sqrt{1-4\alpha^2}\right) - \log 2\alpha}, & (6.38b) \end{cases}$$

where (6.38a) is valid for $R = 1$ and $\alpha_7^* < \alpha < \widehat{\alpha}$ and (6.38b) is valid for $R = 2$.

The functions (6.38a) and (6.38b) are shown in Figure 6.19. Similarly, in the case of $R = 1$, assuming we do not choose $\alpha$ really close to $\alpha_7^*$ then we find that node $N$ is, at worse, ranked 6 positions less than node $L$. In the case when $R = 2$ we see that $p_L(k)_{\max} < 1$ indicating that node $N$ will be ranked more important than node $L - 1$ (or equivalently node $L + 1$).

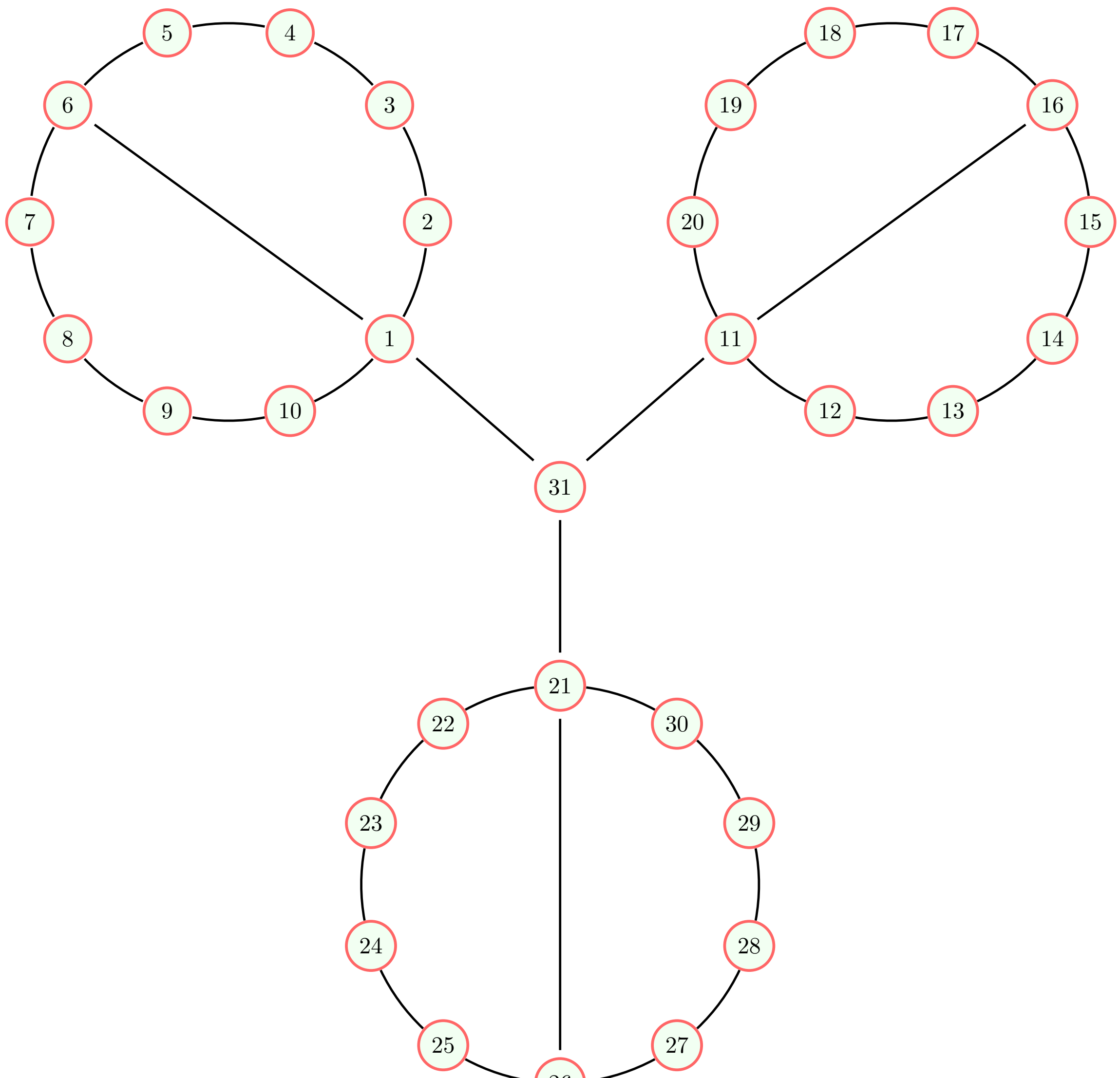
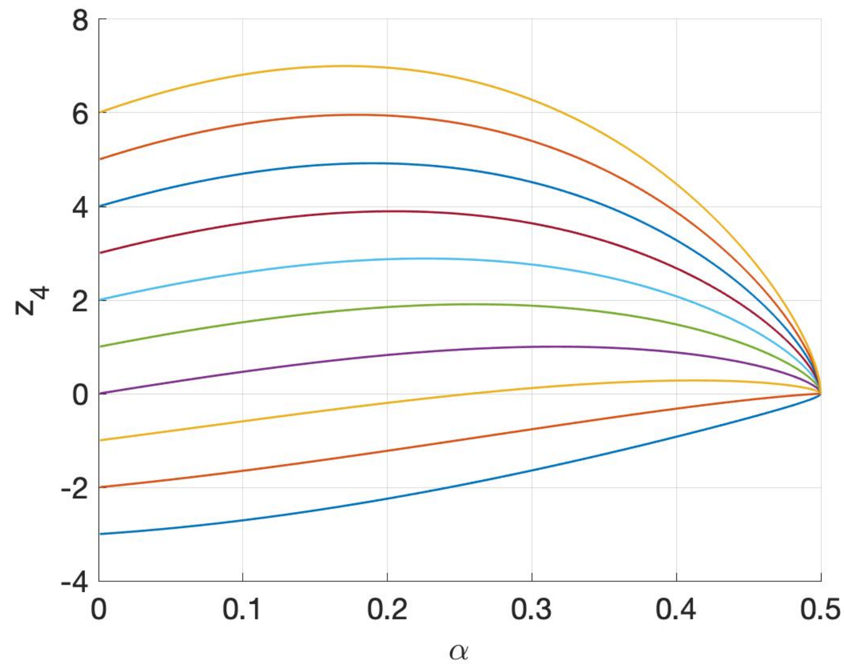
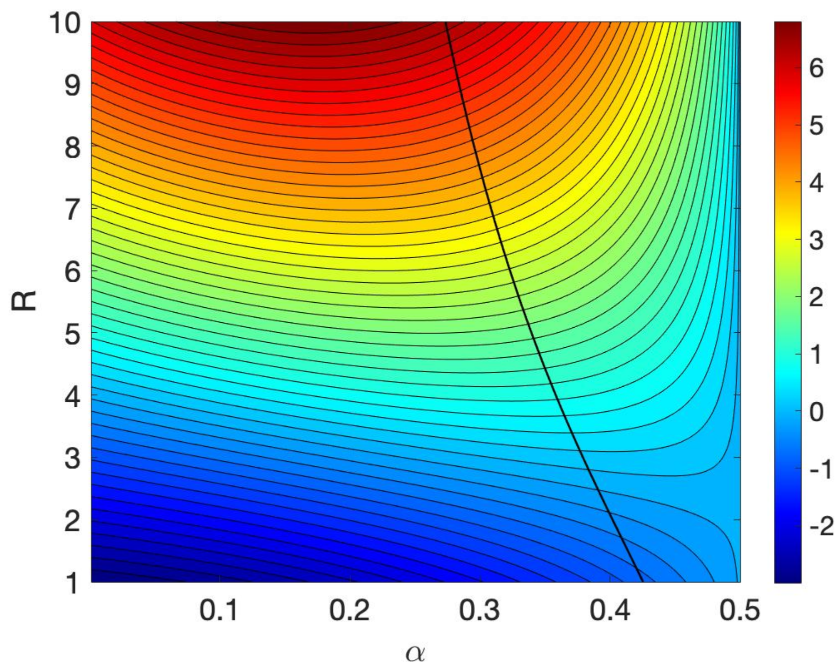



Figure 6.13: *R rings with undirected shortcut and undirected hub*: Illustration with undirected shortcut and undirected hub ($R = 3$, $m = 10$ and $L = 6$).

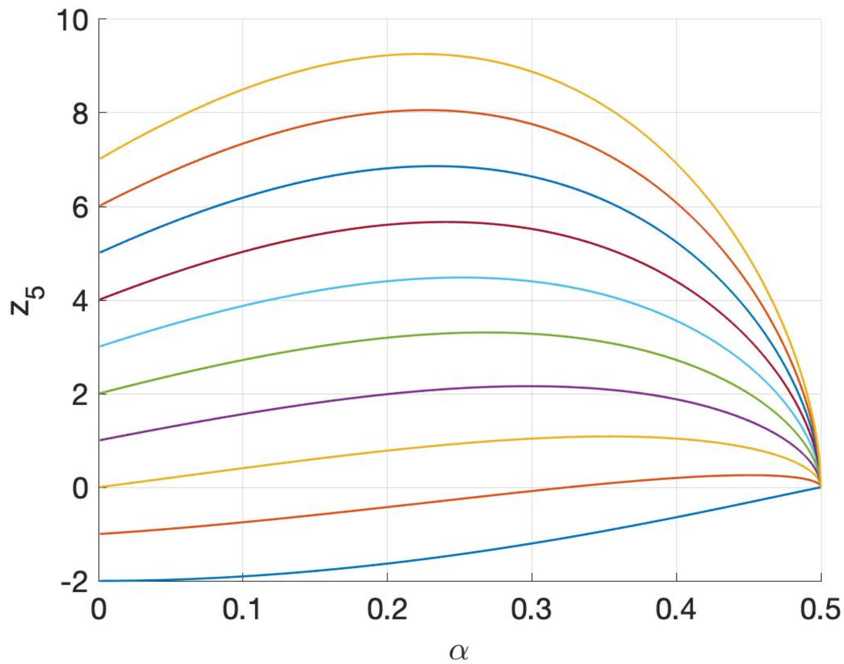



(a) Fixed $R$ ($R$=seq$(1,10)$). $R = 1$ – bottom curve.

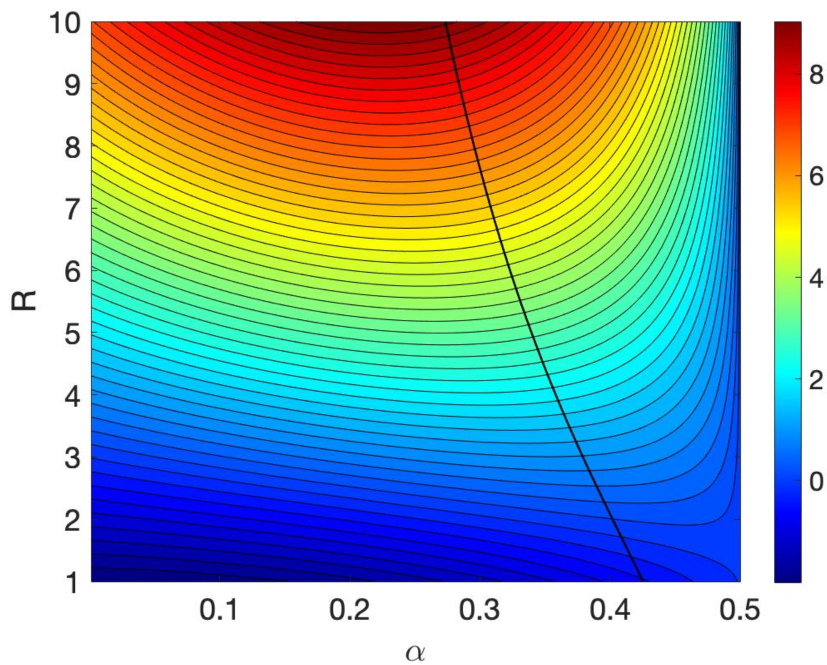



(b) Contour plot of $z_4$ with the solid black line representing $\alpha = \widehat{\alpha}$.

Figure 6.14: *$R$ rings with undirected shortcut and undirected hub*: Relationship between number of rings $R$, Katz parameter $\alpha$ and $z_4$ in (6.34)

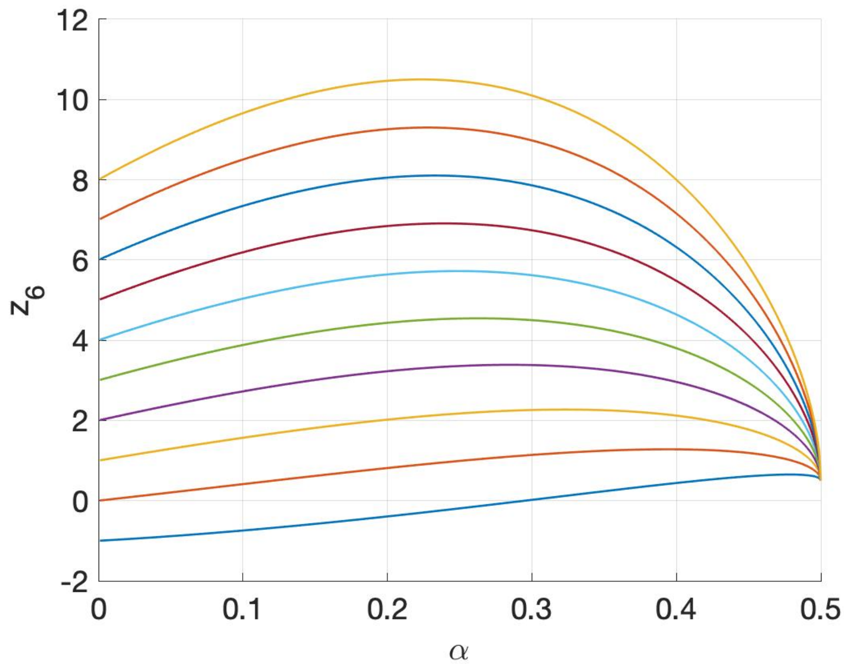



(a) Fixed $R$ ($R$=seq$(1, 10)$). $R = 1$ – bottom curve.

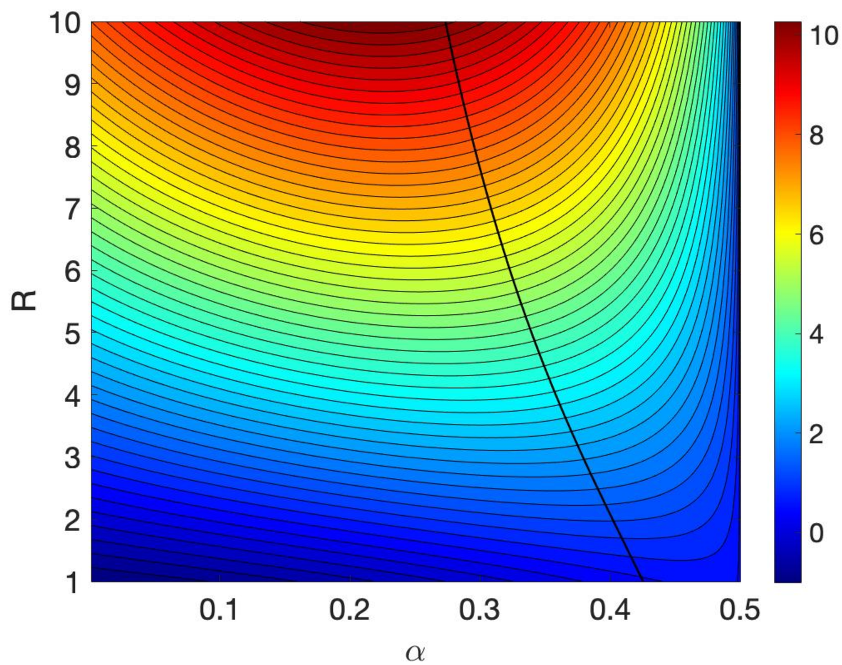



(b) Contour plot of $z_5$ with the solid black line representing $\alpha = \widehat{\alpha}$.

Figure 6.15: *R rings with undirected shortcut and undirected hub*: Relationship between number of rings $R$, Katz parameter $\alpha$ and $z_5$ in (6.35)

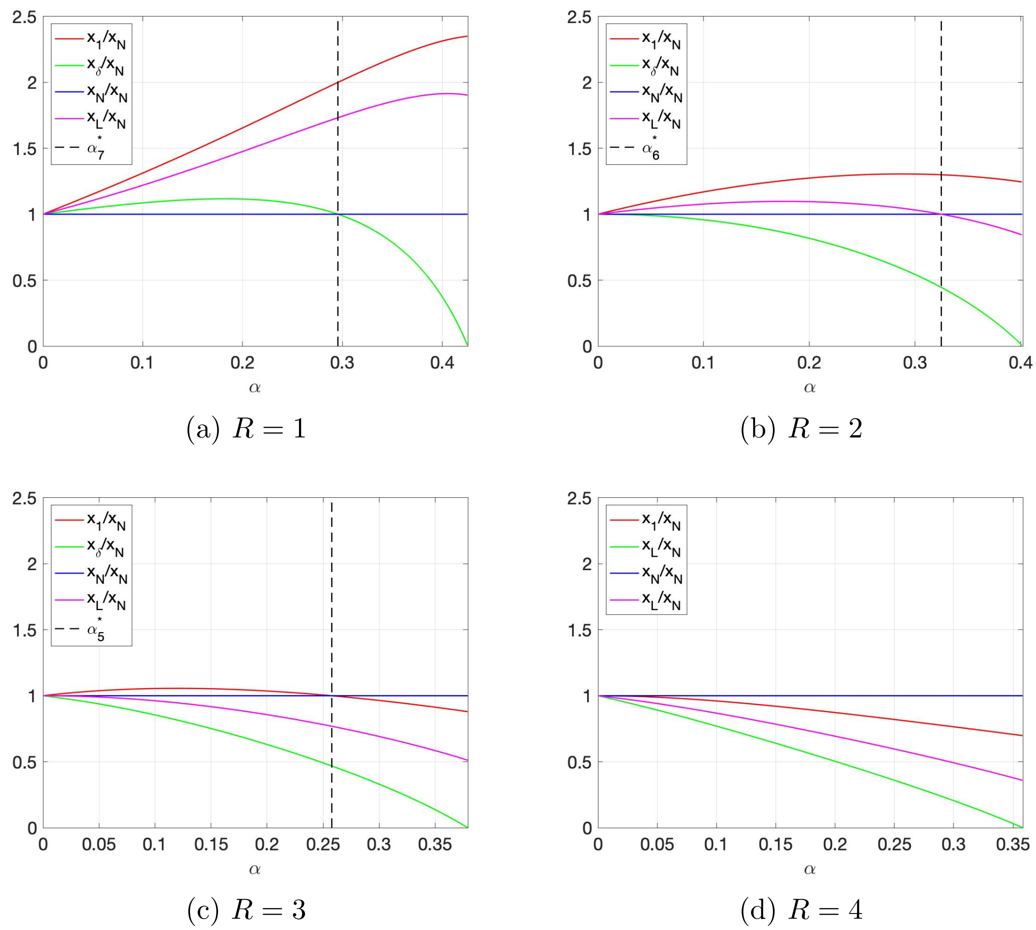



(a) Fixed $R$ ($R$=seq$(1, 10)$). $R = 1$ – bottom curve.




(b) Contour plot of $z_6$ with the solid black line representing $\alpha = \widehat{\alpha}$.

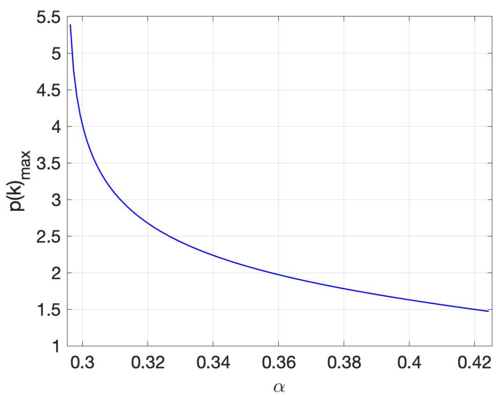
Figure 6.16: *$R$ rings with undirected shortcut and undirected hub*: Relationship between number of rings $R$, Katz parameter $\alpha$ and $z_6$ in (6.36)



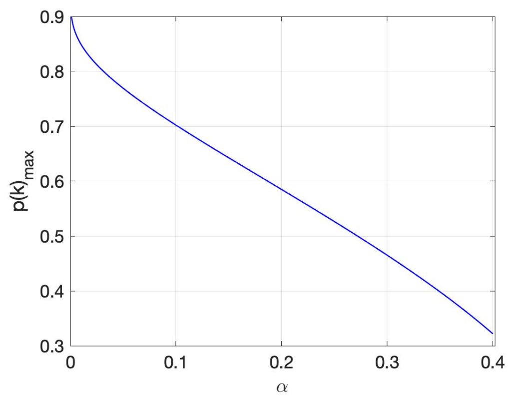

(a) $R = 1$ (b) $R = 2$

(c) $R = 3$ (d) $R = 4$

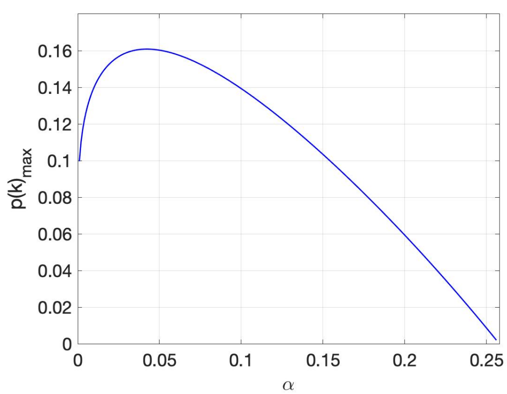
Figure 6.17: *R rings with undirected shortcut and undirected hub*: Ratio of $x_1$ and $x_L$ to $x_N$ for the Katz centrality. We used $m = 1,000$ so $N = 1000R + 1$ with $L = 501$.




(a) $R = 1$

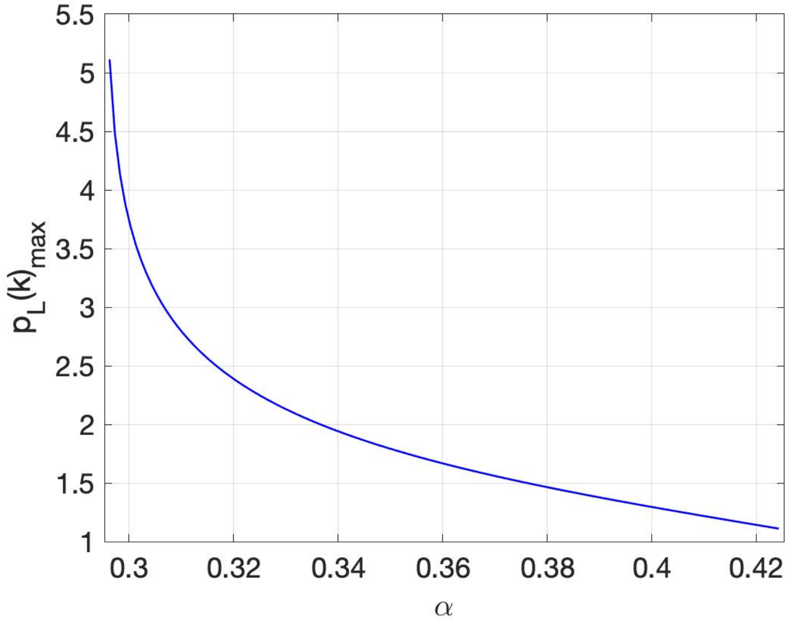



(b) $R = 2$

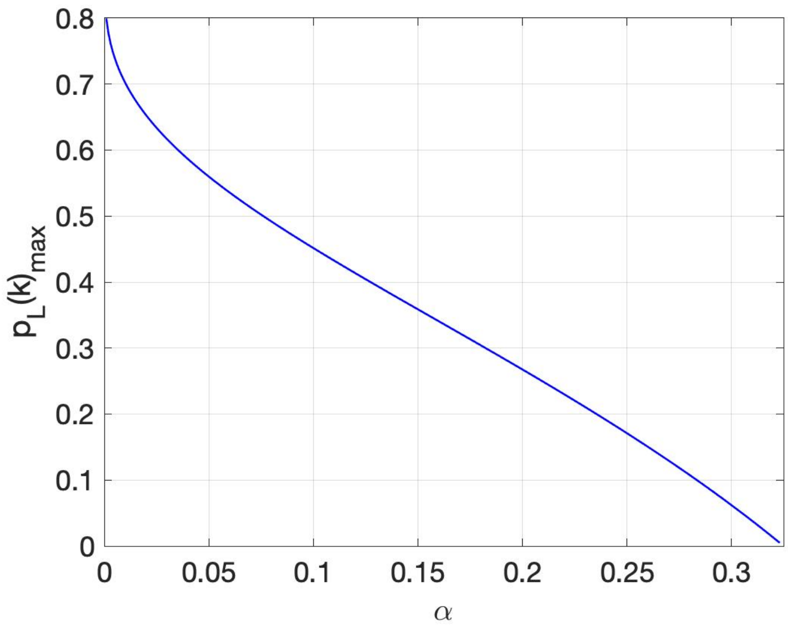



(c) $R = 3$

Figure 6.18: *R rings with undirected shortcut and undirected hub*: $p(k)_{\max}$ indicating the maximum periodic distance for which $x_k > x_N$ for nodes close to node 1




(a) $R = 1$




(b) $R = 2$

Figure 6.19: *R rings with undirected shortcut and undirected hub*: $p_L(k)_{\max}$ indicating the maximum periodic distance for which $x_k > x_N$ for nodes close to node $L$

## 6.3 Summary of key results

In this chapter we have shown that the style of analysis presented in previous chapters can be applied to more complicated networks where results from centrality rankings are not entirely predictable. We have studied a class of network defined by $R$ identical periodic nearest neighbour rings with different configurations of shortcuts and connections to a central hub node.

We used the ansatz from previous chapters as an approximation that we have shown to be highly accurate in a large network limit. This allowed us to find highly accurate closed form expressions which relate the Katz parameter $\alpha$ to the height of the spike resulting from a shortcut across the ring. Requiring the height of the spike to be positive places bounds on the domain for $\alpha$ which intuitively corresponds to that of the Katz system, i.e., $0 < \alpha < 1/\rho(A)$. This led to analytical estimates for $\rho(A)$ that were seen to be extremely accurate.

This analysis also allowed us to characterise the Katz centrality score for nodes in the rings and thus compare key nodes in each network configuration to the central hub node. As Katz centrality is an interpolation of degree centrality and eigenvector centrality we have been able to capture analytically rather than experimentally how Katz centrality morphs between these two different centrality measures. In particular, we found

- Directed shortcut and directed hub:
  Key nodes are 1 (maximum in ring), $\delta$ (minimum in ring) and $N$ (central hub). When $R = 1$ there is a cut off value $\alpha_1^*$ where the rank of nodes $\delta$ and $N$ change.

- Directed shortcut and undirected hub:
  Key nodes are 1 (maximum in ring), $\delta$ (minimum in ring) and $N$ (central hub). When $R = 1$ there is a cut off value $\alpha_3^*$ where the rank of nodes $\delta$ and $N$ change. Similarly, when $R = 3$ there is a cut off value $\alpha_2^*$ where the rank of nodes 1 and $N$ change.

- Undirected shortcut and directed hub:
  Key nodes are 1 (maximum in ring – note node $L$ is equivalent), $\delta$ (minimum

in ring) and $N$ (central hub). When $R = 1$ there is a cut off value $\alpha_4^*$ where the rank of nodes $\delta$ and $N$ change.

- Undirected shortcut and undirected hub:
  Key nodes are 1 (maximum in ring), $L$ (another spike node), $\delta$ (minimum in ring) and $N$ (central hub). There is a cut off value in the case of $R = 1$ where the rank of nodes $\delta$ and $N$ change on passing $\alpha_7^*$. Similarly, there are threshold values when $R = 2$ where rank of nodes $L$ and $N$ change on passing $\alpha_6^*$ and when $R = 3$ there is a threshold value $\alpha_5^*$ beyond which the rank of nodes 1 and $N$ change.

A full summary of rankings can be found in Table 6.9.

Furthermore, in instances where the Katz centrality score of the central hub node was greater than the minimum scored node in the ring but less than the maximum scored node (i.e., $x_1 > x_N > x_\Delta$) we have shown where node $N$ ranks relative to nodal positions in the ring for a given choice of parameter $\alpha$.

These results completely characterise the Katz centrality score and thus the nodal rankings thereby giving insight into the influence of the Katz parameter $\alpha$.

| Network configuration | R | Degree | Katz | Eigenvector |
|---|---|---|---|---|
| Directed shortcut Directed hub | 1 | $x_1 > x_\delta > x_N$ | $x_1 > x_\delta > x_N \ \forall \alpha \in (0, \alpha_1^*)$ | $x_1 > x_N > x_\delta$ |
| | | | $x_1 > x_N > x_\delta \ \forall \alpha \in (\alpha_1^*, \widehat{\alpha})$ | |
| | 2 | $x_1 > x_N = x_\delta$ | $x_1 > x_N > x_\delta \ \forall \alpha \in (0, \widehat{\alpha})$ | $x_1 > x_N > x_\delta$ |
| | 3 | $x_N = x_1 > x_\delta$ | $x_N > x_1 > x_\delta \ \forall \alpha \in (0, \widehat{\alpha})$ | $x_N > x_1 > x_\delta$ |
| | $\geq 4$ | $x_N > x_1 > x_\delta$ | $x_N > x_1 > x_\delta \ \forall \alpha \in (0, \widehat{\alpha})$ | $x_N > x_1 > x_\delta$ |
| Directed shortcut Undirected hub | 1 | $x_1 > x_\delta > x_N$ | $x_1 > x_\delta > x_N \ \forall \alpha \in (0, \alpha_3^*)$ | $x_1 > x_N > x_\delta$ |
| | | | $x_1 > x_N > x_\delta \ \forall \alpha \in (\alpha_3^*, \widehat{\alpha})$ | |
| | 2 | $x_1 > x_N = x_\delta$ | $x_1 > x_N > x_\delta \ \forall \alpha \in (0, \widehat{\alpha})$ | $x_1 > x_N > x_\delta$ |
| | 3 | $x_1 > x_N > x_\delta$ | $x_1 > x_N > x_\delta \ \forall \alpha \in (0, \alpha_2^*)$ | $x_N > x_1 > x_\delta$ |
| | | | $x_N > x_1 > x_\delta \ \forall \alpha \in (\alpha_2^*, \widehat{\alpha})$ | |
| | $\geq 4$ | $x_N \geq x_1 > x_\delta$ | $x_N > x_1 > x_\delta \ \forall \alpha \in (0, \widehat{\alpha})$ | $x_N > x_1 > x_\delta$ |
| Undirected shortcut Directed hub | 1 | $x_1 > x_\delta > x_N$ | $x_1 > x_\delta > x_N \ \forall \alpha \in (0, \alpha_4^*)$ | $x_1 > x_N > x_\delta$ |
| | | | $x_1 > x_N > x_\delta \ \forall \alpha \in (\alpha_4^*, \widehat{\alpha})$ | |
| | 2 | $x_1 > x_N = x_\delta$ | $x_1 > x_N > x_\delta \ \forall \alpha \in (0, \widehat{\alpha})$ | $x_1 > x_N > x_\delta$ |
| | 3 | $x_N = x_1 > x_\delta$ | $x_N > x_1 > x_\delta \ \forall \alpha \in (0, \widehat{\alpha})$ | $x_N > x_1 > x_\delta$ |
| | $\geq 4$ | $x_N > x_1 > x_\delta$ | $x_N > x_1 > x_\delta \ \forall \alpha \in (0, \widehat{\alpha})$ | $x_N > x_1 > x_\delta$ |
| Undirected shortcut Undirected hub | 1 | $x_1 > x_L > x_\delta > x_N$ | $x_1 > x_L > x_\delta > x_N \ \forall \alpha \in (0, \alpha_7^*)$ | $x_1 > x_L > x_N > x_\delta$ |
| | | | $x_1 > x_L > x_N > x_\delta \ \forall \alpha \in (\alpha_7^*, \widehat{\alpha})$ | |
| | 2 | $x_1 > x_L > x_N = x_\delta$ | $x_1 > x_L > x_N > x_\delta \ \forall \alpha \in (0, \alpha_6^*)$ | $x_1 > x_N > x_L > x_\delta$ |
| | | | $x_1 > x_N > x_L > x_\delta \ \forall \alpha \in (\alpha_6^*, \widehat{\alpha})$ | |
| | 3 | $x_1 > x_N = x_L > x_\delta$ | $x_1 > x_N > x_L > x_\delta \ \forall \alpha \in (0, \alpha_5^*)$ | $x_N > x_1 > x_L > x_\delta$ |
| | | | $x_N > x_1 > x_L > x_\delta \ \forall \alpha \in (\alpha_5^*, \widehat{\alpha})$ | |
| | $\geq 4$ | $x_N \geq x_1 > x_L > x_\delta$ | $x_N > x_1 > x_L > x_\delta \ \forall \alpha \in (0, \widehat{\alpha})$ | $x_N > x_1 > x_L > x_\delta$ |

Table 6.9: Katz centrality for key nodes in the various different network class configurations examined in this Chapter.

# Chapter 7

# PageRank centrality

In this Chapter we apply a similar style of analysis as Chapter 4 but this time we examine PageRank centrality which has a different feel to that of Katz and eigenvector centrality in that it's concerned with the ability to receive information as opposed to send information. We finish this Chapter by putting our result into use by examining a more complicated network where the nodal rankings are not straightforward.

## 7.1 Exact solution for PageRank centrality

We now consider the PageRank system (3.5) on a periodic ring plus a directed shortcut. To be consistent with the treatment in sections 4.3 and 4.4, we will ensure that node 1 remains the most highly ranked. So the directed shortcut will be added *from node $L$ to node* 1. Hence, the adjacency matrix now has the form $A = C + \widehat{E}$, where the rank one matrix $\widehat{E}$ is zero except for $E(L, 1) = 1$.

Figure 7.1 relates to the case where $N = 20$, $L = 8$ and $d = 0.8$. Asterisks in the upper picture show the components of the PageRank vector $x$ in (3.5). We see that node 1 is ranked highest, and node $L$, which gives away the extra shortcut, has a low ranking that is slightly higher than its two neighbours. Unlike the cases shown in Figures 4.2 and 4.3, the solution does not appear to be periodically symmetric about node 1.

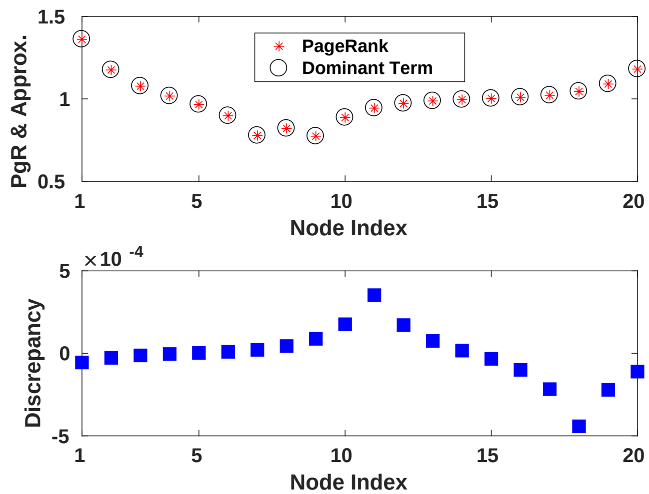



Figure 7.1: Upper picture: asterisks show components of the PageRank vector $x$ from (3.5) and circles show the periodic spike approximation (7.3)–(7.4). Lower picture: the discrepancy between these two vectors.

Figure 7.2 gives a view of the asymptotic $N \to \infty$ structure by increasing $N$ to 100, with $L = 40$. The solution now appears to be periodically *decreasing* locally away from node 1 and periodically *increasing* locally away from node $L$.

Returning to the $N = 20$ ring, the circles in the upper picture of Figure 7.1 show the approximation arising from fitting suitable spikes. The lower picture in Figure 7.1 shows the discrepancy, which takes relatively small values that peak and trough at the nodes diametrically opposite 1 and $L$, respectively. We also see in Figures 7.1 and 7.2 that node $L$ does not quite fit into the general pattern of periodic growth/decay.

The solutions seen in Figures 7.1 and 7.2 may be likened to a positive spike centered at node 1 plus a negative spike centered at node $L$, so we may expect $x_i$ to contain a term proportional to $u_1^{p(i)}$ and a term proportional to $-u_1^{p_L(i)}$, where we recall that $p_L(i)$ in (5.2) denotes the periodic distance from node $i$ to node $L$.

An exception occurs at node $L$, which seems to break this pattern. Theorem 7.1 below shows that the solution does indeed have this general form—plus exponentially small terms—with a simple shift needed at node $L$. The general form (7.1)–(7.2) involves five constants $g_1, g_1', g_2, g_2', f$ and our proof technique relies

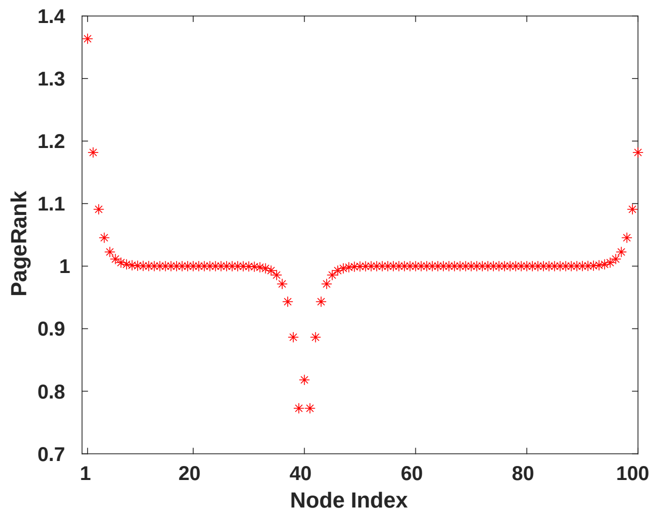



Figure 7.2: Asterisks show components of the PageRank vector $x$ from (3.5) for a larger network: here $N = 100$ and $L = 40$.

on five "special" nodes that impose independent constraints. For this reason, we need to rule out the exceptional case where node $L$ is within one hop of being diametrically opposite node 1, which would lead to fewer than five special nodes. We therefore assume that the shortcut is chosen so that $0 < \theta < 1$ in (4.4). We also treat the PageRank parameter $0 < d < 1$ as being fixed, independently of $N$.

**Theorem 7.1.** *For the undirected ring plus directed shortcut network with adjacency matrix $A = C + \widehat{E}$, for $0 < \theta < 1$ and for sufficiently large $N$ the unique solution of the PageRank system (3.5) has the form*

$$x_i = 1 + g_1 u_1^{p(i)} + g_2 u_2^{p(i)} + g_1' u_1^{p_L(i)} + g_2' u_2^{p_L(i)}, \quad \text{for } i \neq L, \tag{7.1}$$

*and*

$$x_L = 1 + g_1 u_1^{L-1} + g_2 u_2^{L-1} + g_1' + g_2' + f. \tag{7.2}$$

*Here, $u_1, u_2, g_1, g_2, g_1', g_2', f$ are constants, i.e., independent of $i$, and $u_1, u_2$ are independent of $N$. In particular, $g_1 > 0$, $g_2 > 0$, $g_1' < 0$, $g_2' < 0$, $f > 0$, and*

*$u_1, u_2$ are the roots of the palindromic quadratic $du^2 - 2u + d$, so that*

$$u_1 = \frac{1 - \sqrt{1-d^2}}{d}, \quad u_2 = \frac{1 + \sqrt{1-d^2}}{d}.$$

*Hence, $u_2 = 1/u_1$ and $0 < u_1 < 1 < u_2$. Moreover, the terms in (7.1) and (7.2) involving $g_2$ and $g_2'$ are exponentially small asymptotically, in the sense that*

$$x_i = 1 + g_1 u_1^{p(i)} + g_1' u_1^{p_L(i)} + O(u_1^{N/2}), \quad \text{for } i \neq L, \tag{7.3}$$

$$x_L = 1 + g_1' + f + O(u_1^{L-1}). \tag{7.4}$$

*We also have*

$$g_1 = \frac{d}{1 + 2\sqrt{1-d^2}} + O(u_1^{\theta N/2}), \tag{7.5}$$

$$g_1' = \frac{-1}{1 + 2\sqrt{1-d^2}} + O(u_1^{\theta N/2}), \tag{7.6}$$

$$f = \frac{\sqrt{1-d^2}}{1 + 2\sqrt{1-d^2}} + O(u_1^{\theta N/2}). \tag{7.7}$$

*Proof.* As in the proofs of Theorems 4.1 and 4.2, we show by direct substitution that $x$ satisfies the required conditions. Because the style of analysis is similar, we omit some of the fine details.

We begin with the ansatz

$$x_i = b + gu^{p(i)} + g'u^{p_L(i)} + \epsilon f, \tag{7.8}$$

where $\epsilon = 1$ if $i = L$ and $\epsilon = 0$ for $i \neq L$.

For a general node $k$, where each neighbouring node differs by a factor of $u$, the PageRank system (3.5) requires

$$x_k - \frac{d}{2}(x_{k-1} + x_{k-1}) = 1 - d. \tag{7.9}$$

Using (7.8) and substituting into (7.9) we have

$$b(1-d) + g_1 u^{p(j)} \left(1 - \frac{1}{2} d \left(u + u^{-1}\right)\right) + g' u^{p_L(j)} \left(1 - \frac{1}{2} d \left(u + u^{-1}\right)\right) = 1 - d. \tag{7.10}$$

We can satisfy this equation independently of $k$ by setting $b = 1$ and choosing $u$ to be either root of the quadratic $du^2 - 2u + d$. By linearity of (7.9) we may extend (7.8) to include a linear combination involving both roots, so our ansatz becomes

$$x_i = 1 + g_1 u^{p(i)} + g_2 u^{p(i)} + g_1' u^{p_L(i)} + g_2' u^{p_L(i)} + \epsilon f, \tag{7.11}$$

where

$$u_1 = \frac{1 - \sqrt{1 - d^2}}{d}, \quad u_2 = \frac{1 + \sqrt{1 - d^2}}{d}.$$

Once again, since the quadratic is palindromic, $u_1$ ad $u_2$ must satisfy $u_1 u_2 = 1$. It is clear that these roots are real with $0 < u_1 < 1 < u_2$ and we note the useful facts

$$1 - d\left(\frac{1}{3} u_1^{-1} + \frac{1}{2} u_1\right) = \frac{1}{6} d u_1^{-1}, \qquad 1 - d\left(\frac{1}{3} u_2^{-1} + \frac{1}{2} u_2\right) = \frac{1}{6} d u_2^{-1}. \tag{7.12}$$

It remains to check the system at

- node 1, where there are three incoming edges,
- node $L$, where, from (7.2), the general form of the solution has undergone a shift,
- node $L + 1$ (or, equivalently, node $L - 1$), where the nonstandard node $L$ is involved,
- node $N/2 + 1$, which has neighbours that are the same periodic distance from node 1, (assuming $N$ is even). If $N$ is odd we would consider either node $(N + 1)/2$ or $(N + 3)/2$, where there is a single neigbour at the same periodic distance.
- node $N/2 + L$, which has neighbours that are the same periodic distance from node $L$, (assuming $N$ is even). If $N$ is odd we would consider either

node $(N-1)/2+L$ or $(N+1)/2+L$, where there is a single neigbour at the same periodic distance.

Our equation for node 1 is $x_1 - d(x_2/2 + x_N/2 + x_L/3) = 1 - d$. Using (7.1) and (7.2), and simplifying, we arrive at

$$g_1\left(1 - d\left(u_1 + \tfrac{1}{3}u_1^{L-1}\right)\right) + g_2\left(1 - d\left(u_2 + \tfrac{1}{3}u_2^{L-1}\right)\right) - \tfrac{1}{3}d\left(g_1' + g_2' + f + 1\right) = 0. \tag{7.13}$$

For node $L$, we have $x_L - d(x_{L-1}/2 + x_{L+1}/2) = 1 - d$, which simplifies to

$$g_1'(1 - du_1) + g_2'(1 - du_2) + f = 0. \tag{7.14}$$

Node $L+1$ gives $x_{L+1} - d(x_L/3 + x_{L+2}/2) = 1 - d$, from which, using (7.12), we obtain

$$g_1 u_1^{L-1} + g_2 u_2^{L-1} + g_1' + g_2' - 2f = -1. \tag{7.15}$$

At node $N/2+1$ we have $x_{N/2+1} - d(x_{N/2}/2 + x_{N/2+2}/2) = 1 - d$, which leads to

$$g_2 = g_1 u_1^N. \tag{7.16}$$

Similarly, at node $N/2+L$, the condition $x_{N/2+L} - d(x_{N/2+L-1}/2 + x_{N/2+L+1}/2) = 1 - d$ leads to

$$g_2' = g_1' u_1^N. \tag{7.17}$$

In the derivation of (7.16) and (7.17) we have assumed $N$ to be even. The same result holds in the case of $N$ odd.

Using (7.16) and (7.17) to eliminate $g_2$ and $g_2'$, we are left with three linear equations, (7.13)–(7.15), for the three unknowns $g_1$, $g_1'$ and $f$. These may be written in the form

$$(B + \Delta B)\, v = r, \tag{7.18}$$

where

$$B = \begin{bmatrix} 1-du_1 & -d/3 & -d/3 \\ 0 & 1-du_1 & 1 \\ 0 & 1 & -2 \end{bmatrix}, \quad v = \begin{bmatrix} g_1 \\ g_1' \\ f \end{bmatrix}, \quad r = \begin{bmatrix} d/3 \\ 0 \\ -1 \end{bmatrix},$$

and $\Delta B \in \mathbb{R}^{3\times 3}$ is such that $\|\Delta B\|_\infty = O(u_1^{\theta N/2})$. Writing $u_1$ in terms of $d$, the determinant of $B$ reduces to $-(1+2\sqrt{1-d^2})\sqrt{1-d^2} \neq 0$, so $B$ is nonsingular.

From (7.18), we have

$$\left(I + B^{-1}\Delta B\right) v = B^{-1} r.$$

The matrix $I + B^{-1}\Delta B$ is invertible for sufficiently large $N$, which establishes that (7.18) has a unique solution. We conclude that $x$ in (7.1)–(7.2) solves the PageRank system (3.5). Moreover, since $(I + B^{-1}\Delta B)^{-1} = I + O(\Delta B)$, we see that $v = B^{-1}r + O(u_1^{\theta N/2})$, which leads to (7.5)–(7.7). The relations (7.16) and (7.17) then show that $g_2 > 0$ and $g_2' < 0$ and also give the expansions (7.3) and (7.4).

In a similar manner to previous theorems, we know from (7.16) and (7.17) that $g_2 = O(u_1^N)$ and $g_2' = O(u_1^N)$. Since $g_2 u_2^{p(k)} = g_1 u_1^{N-p(k)}$ and $g_2' u_2^{p_L(k)} = g_1' u_1^{N-p_L(k)}$ is largest when $k = N/2+1$ and $k = N/2+L$ respectively. Therefore for $i \neq L$

$$\begin{aligned} x_i &= 1 + g_1 u_1^{p(i)} + g_2 u_2^{p(i)} + g_1' u_1^{p_L(i)} + g_2' u_2^{p_L(i)} \\ &= 1 + g_1 u_1^{p(i)} + g_1 u_1^{N-p(i)} + g_1' u_1^{p_L(i)} + g_1' u_1^{N-p_L(i)} \\ &= 1 + g_1 u_1^{p(i)} + g_1' u_1^{p_L(i)} + O(u_1^{N/2}). \end{aligned}$$

When considering node $L$ specifically (due to the shift), we have

$$\begin{aligned} x_L &= 1 + g_1 u_1^{L-1} + g_2 u_2^{L-1} + g_1' + g_2' + f \\ &= 1 + g_1 u_1^{L-1} + g_1 u_1^{N-L+1} + g_1' + g_1' u_1^N + f \\ &= 1 + g_1' + f + O(u_1^{L-1}). \end{aligned}$$

□

For the examples in Figures 7.1 and 7.2, we see that the node giving out the extra link, with index $L$, has a slightly larger PageRank value than its immediate neighbours, but not more than its second-neighbours. The following corollary shows that this behaviour is generic.

**Corollary 7.1.** *For the undirected ring plus directed shortcut network with adjacency matrix $A = C + \widehat{E}$, for sufficiently large $N$ the PageRank vector satisfies*

$$\max\left(x_{L+1}, x_{L-1}\right) < x_L < \min\left(x_{L+2}, x_{L-2}\right).$$

*Proof.* It follows from Theorem 7.1 that, ignoring asymptotically small terms,

$$x_L = 1 + g_1' + f, \quad x_{L+1} = x_{L-1} = 1 + g_1' u_1, \quad x_{L+2} = x_{L-2} = 1 + g_1' u_1^2.$$

Using (7.5), (7.6) and (7.7) we find that

$$x_L - x_{L+1} = x_L - x_{L-1} = \frac{(1-d)\left(1-\sqrt{1-d^2}\right)}{d(1+2\sqrt{1-d^2})} > 0,$$

and

$$x_L - x_{L+2} = x_L - x_{L-2} = \frac{-\sqrt{1-d^2}\left(\sqrt{1-d^2}-1\right)^2}{d^2(1+2\sqrt{1-d^2})} < 0.$$

□

We can at least partially explain the ordering in Corollary 7.1 by noting that nodes $L+1$ and $L-1$ suffer because one of their two neighbours, node $L$, has three outgoing links. Hence the "vote of confidence" from node $L$ is split three ways instead of two. As a consequence, node $L$ itself suffers from having two lowly-ranked neighbours. What is perhaps less obvious, but proved in Corollary 7.1, is that nodes $L-1$ and $L+1$ can never rise above node $L$, which itself can never rise above nodes $L-2$ and $L+2$.

## 7.2 Example network

In this section we consider the case of each ring in the network having a directed shortcut across the ring. As each ring is identical, we denote this using ring 1 as an edge from node $L$ to node 1 where $1 < L \leq m/2 + 1$.

We consider the case of each of the $R$ rings connecting to the central hub node with an outgoing directed edge from each of the local node 1s to the central hub node $N$. Figure 7.3 illustrates an example network with $R = 3$ rings where each ring contains the same identical directed shortcut from local node 6 to local node 1 (i.e. $6 \mapsto 1, 16 \mapsto 11, 26 \mapsto 21$). Each local node 1 is connected to the central node 31 via an outgoing directed edge (i.e. $1 \mapsto 31, 11 \mapsto 31, 21 \mapsto 31$).

We explore how the influence of the PageRank parameter $d$ effects the ranking of the central node N.

For the general $R$-ring network, the PageRank system (3.5) reduces to

$$x_k - d\left(\frac{1}{2}x_{k-1} + \frac{1}{2}x_{k+1}\right) = 1 - d, \tag{7.19}$$

$$x_1 - d\left(\frac{1}{2}x_2 + \frac{1}{2}x_m + \frac{1}{3}x_L\right) = 1 - d, \tag{7.20}$$

$$x_2 - d\left(\frac{1}{3}x_1 + \frac{1}{2}x_3\right) = 1 - d, \tag{7.21}$$

$$x_m - d\left(\frac{1}{2}x_{m-1} + \frac{1}{3}x_1\right) = 1 - d, \tag{7.22}$$

$$x_L - d\left(\frac{1}{2}x_{L-1} + \frac{1}{2}x_{L+1}\right) = 1 - d, \tag{7.23}$$

$$x_{L-1} - d\left(\frac{1}{2}x_{L-2} + \frac{1}{3}x_L\right) = 1 - d, \tag{7.24}$$

$$x_{L+1} - d\left(\frac{1}{3}x_L + \frac{1}{2}x_{L+2}\right) = 1 - d, \tag{7.25}$$

$$x_N - \frac{1}{3}Rdx_1 = 1 - d. \tag{7.26}$$

The general equation (7.19) is valid for $3 \leq k \leq L - 2$ and $L + 2 \leq k \leq m - 1$. Whilst the form given by (7.19) when $k = L$ is correct we have listed this specifically as (7.23) as node $L$ has a shift that we will need to quantify for our analysis.

We know from Theorem 7.1, with the shortcut in the rings, that we should observe growth at node 1, decay at node $L$ and a shift at node $L$. In this instance, we have an additional edge originating from each of the local node 1s to the central hub node $N$ and therefore there will be decay at node 1 and a shift at node 1. We therefore have growth and decay at node 1, the net effect of which will need to be quantified, which can be combined into a single term in our ansatz. Let

$$x_i = b + gu^{p(i)} + g'u^{p_L(i)} + \epsilon_1 f_1 + \epsilon_L f_L, \tag{7.27}$$

for the centrality of nodes in the ring where $\epsilon_1 = 1$ if $i = 1$ and 0 otherwise, $\epsilon_L = 1$ if $i = L$ and 0 otherwise.

Inserting (7.27) we find that (7.19) is solved with $b = 1$ and $u = u_1$. Focusing only on the dominant parts of the PageRank vector it follows that we need only consider one neighbour of nodes 1 and $L$, i.e. (7.21) is equivalent to (7.22) and (7.24) is equivalent to (7.25).

Using (7.27) in (7.20), (7.21), (7.23) and (7.25) we have

$$\begin{aligned} g(1-du) + f_1 - \frac{1}{3}d(1 + g' + f_L) &= 0, &(7.28)\\ g - 2f_1 &= -1, &(7.29)\\ g'(1-du) + f_L &= 0, &(7.30)\\ g' - 2f_L &= -1. &(7.31)\end{aligned}$$

Solving the system (7.28) – (7.31) leads to the solutions

$$\begin{aligned} g &\approx \frac{2(d-1)\sqrt{1-d^2} - 1}{[1 + 2\sqrt{1-d^2}]^2} < 0, &(7.32)\\ g' &\approx -\frac{1}{1 + 2\sqrt{1-d^2}}, < 0 &(7.33)\\ f_1 &\approx \frac{2(d-1)\sqrt{1-d^2} - 1}{2[1 + 2\sqrt{1-d^2}]^2} + \frac{1}{2} < 0, &(7.34)\\ f_L &\approx \frac{\sqrt{1-d^2}}{1 + 2\sqrt{1-d^2}} > 0. &(7.35)\end{aligned}$$

From (7.32) we know that $g < 0$ and therefore the effect of the growth and decay at node 1 leads to a net negative decay and so a negative spike will be observable around node 1. Nodes which are not close to 1 or $L$ will obtain a centrality score of 1. We need to know if node 1 can attain a better nodal ranking due to the shift $f_1$, i.e. is $x_1 = 1 + g + f_1 > 1$. Using (7.32) and (7.34) this leads to

$$\frac{6(d-1)\sqrt{1-d^2}-3}{2[1+2\sqrt{1-d^2}]^2} + \frac{1}{2} > 0, \tag{7.36}$$

which when multiplied by a factor of $2[1+2\sqrt{1-d^2}]^2$ and rearranged reduces to the condition

$$(3d-1)\sqrt{1-d^2} + 1 - 2d^2 > 0. \tag{7.37}$$

Solving for the roots in our range $0 < d < 1$ gives a cut-off value of

$$d_1^* \approx 0.9195, \tag{7.38}$$

for which this condition is met. Therefore, if $0 < d < d_1^*$ we have $x_1 > 1$ and if $d_1^* < d < 1$ then $x_1 < 1$.

Due to the shifts at nodes 1 and $L$ we hereby characterise the ranking of nodes that are immediate neighbours to these nodes, similar to Corollary 7.1. In this instance

$$x_1 = 1 + g + f_1, \qquad x_2 = x_m = 1 + gu_1, \qquad x_3 = x_{m-1} = 1 + gu_1^2,$$

and

$$x_L = 1 + g' + f_L, \qquad x_{L+1} = x_{L-1} = 1 + g'u_1, \qquad x_{L+2} = x_{L-2} = 1 + g'u_1^2.$$

The upper plot of Figure 7.4 shows the value of $x_3 - x_2 = x_{m-1} - x_m$ and $x_3 - x_1 = x_{m-1} - x_1$. As by definition, we expected $x_3 - x_2 > 0$ but what is clear is that since $x_3 - x_1 < 0$ we know that node 1 still manages to attain a higher PageRank centrality value (due to the shift). The lower plot of Figure 7.4 shows the value of $x_L - x_{L+1} = x_L - x_{L-1}$ and $x_L - x_{L+2} = x_L - x_{L-2}$. As previously seen we observe that $x_L - x_{L+1} > 0$ and $x_L - x_{L+2} < 0$ meaning that $x_L$ is sandwiched between $x_{L+1}$ and $x_{L+2}$.

Now that we know the behaviour around the spikes we compare node $N$ to key nodes in the ring. In considering this ring we expect node 1 to be ranked first when $0 < d < d_1^*$. We define nodes sufficiently far away from nodes 1 and $L$, i.e. where the effects of the spikes are negligible, by $\Delta$. In the case when $d_1^* < d < 1$ then we expect node $\Delta$ to be the maximum ranked node.

We know from (7.26) that $x_N = 1 - d + Rd(1 + g + f_1)/3$ so we now turn our attention to when $x_N > x_1$, i.e.

$$3(1-d) + (Rd-3)(1+g+f_1) > 0.$$

Using the expression for $g$ and $f_1$ in (7.32) and (7.34) this gives

$$3(1-d) + (Rd-3)\left[\frac{6(d-1)\sqrt{1-d^2}-3}{2[1+2\sqrt{1-d^2}]^2} + \frac{3}{2}\right] > 0,$$

which when multiplied by a factor of $2[1+2\sqrt{1-d^2}]^2$ can be rearranged to

$$\left[1+2\sqrt{1-d^2}\right]^2[6(1-d)+3(Rd-3)] + 6(Rd-3)(d-1)\sqrt{1-d^2} - 3(Rd-3) > 0. \tag{7.39}$$

After some simplification we arrive at the condition

$$z_7 := \sqrt{1-d^2}\left[Rd^2 + (R-7)d + 1\right] - 2(R-2)d^3 + 2d^2 + (2R-5)d - 1 > 0. \tag{7.40}$$

From Figure 7.5 we observe that $z_7 < 0$ for $R \leq 2$ and $z_7 > 0$ for $R \geq 4$. We therefore conclude that $x_1 > x_N$ when we have 2 or fewer rings and $x_N > x_1$ when we have 4 or more rings. When $R = 3$ there is a pole where $z_7$ changes sign and from (7.40) we have

$$\begin{aligned} z_7 &= \sqrt{1-d^2}\left[3d^2 - 4d + 1\right] - 2d^3 + 2d^2 + d - 1 \\ &= (d-1)\left[(3d-1)\sqrt{1-d^2} + 1 - 2d^2\right] \end{aligned}$$

Since we are working with a multiple of (7.37) it follows that the the $d_1^*$ is the threshold value beyond which node $N$ is ranked higher than node 1, i.e. where $x_N > x_1$.

Within the ring we observe an inverted spike at node $L$, however $x_L$ also contains a shift ($f_L$) and is therefore not the minimum ranked node. The node ranked last, $\delta$, in this instance is node $L-1$ or $L+1$. Using $x_\delta = 1 + g'u_1$ we find $x_N > x_\delta$ when

$$1 - d + \frac{1}{3}Rd(1 + g + f_1) > 1 + g'u_1.$$

Using the expressions for $g, g'$ and $f_1$ in (7.32), (7.33) and (7.34) respectively this gives

$$-d + Rd\left[\frac{2(d-1)\sqrt{1-d^2} - 1}{2[1 + 2\sqrt{1-d^2}]^2} + \frac{1}{2}\right] > -\frac{1 - \sqrt{1-d^2}}{d(1 + 2\sqrt{1-d^2})},$$

which when multiplied by a factor of $2d[1 + 2\sqrt{1-d^2}]^2$ can be rearranged to

$$\begin{aligned}\left[1 + 2\sqrt{1-d^2}\right]^2 d^2(R-2) + 2Rd^2(d-1)\sqrt{1-d^2} - Rd^2 \\ +2(1 - \sqrt{1-d^2})(1 + 2\sqrt{1-d^2}) > 0.\end{aligned} \tag{7.41}$$

Following some simplification we arrive at

$$z_8 := \sqrt{1-d^2}\left[Rd^3 + (R-4)d^2 + 1\right] + (2R-3)d^2 - 2(R-2)d^4 - 1 > 0. \tag{7.42}$$

From Figure 7.6 we observe that when $z_8 < 0$ for $R = 1$ and $z_8 > 0$ for $R \geq 3$ meaning that $x_\delta > x_N$ when we have 1 ring in the network and $x_N > x_\delta$ when there are 3 or more rings. When $R = 2$ there is a pole where $z_8$ changes sign from negative to positive indicating a threshold value beyond which node $N$ moves up the nodal rankings. Solving for the roots of $\sqrt{1-d^2}[2d^3 - 2d^2 + 1] + d^2 - 1$ gives

$$d_2^* \approx 0.7071.$$

So far, we have examined how the ranking of node $N$ compares to that of nodes 1 and $\delta$. For completeness, before we examine the rest of the nodes in the ring, we first pay particular attention to node $L$ since $x_L$ contains the shift $f_L$. Using $x_L = 1 + g' + f_L$ we have $x_N > x_L$ when

$$1 - d + \frac{1}{3}Rd(1 + g + f_1) > 1 + g' + f_L.$$

Using (7.32), (7.33), (7.34) and (7.35) this gives the condition

$$-d + Rd\left[\frac{2(d-1)\sqrt{1-d^2}-1}{2[1+2\sqrt{1-d^2}]^2}+\frac{1}{2}\right] > -\frac{1-\sqrt{1-d^2}}{1+2\sqrt{1-d^2}},$$

which when multiplied by a factor of $2[1+2\sqrt{1-d^2}]^2$ can be rearranged to

$$\begin{aligned}\left[1+2\sqrt{1-d^2}\right]^2 d(R-2) + 2Rd(d-1)\sqrt{1-d^2} - Rd\\ +2(1-\sqrt{1-d^2})(1+2\sqrt{1-d^2}) > 0.\end{aligned} \tag{7.43}$$

Following some simplification we arrive at

$$z_9 := \sqrt{1-d^2}\left[Rd^2+(R-4)d+1\right] - 2(R-2)d^3 + 2d^2 + (2R-5)d - 1 > 0. \tag{7.44}$$

From Figure 7.7 we observe that when $z_9 < 0$ for $R = 1$ and $z_9 > 0$ for $R \geq 3$ meaning that $x_L > x_N$ when we have 1 ring in the network and $x_N > x_L$ when there are 3 or more rings. When $R = 2$ there is a pole where $z_9$ changes sign from negative to positive indicating a threshold value beyond which node $N$ becomes more central than node $L$. Solving for the roots of $\sqrt{1-d^2}[2d^2-2d+1]+2d^2-d-1$ gives

$$d_3^* \approx 0.8565.$$

Thus far we have examined how node $N$ compares to nodes $1, L$ and $\delta$. In the instance when $d^* < d < 1$ node 1 is not the highest ranked node. Instead this is, as defined, node $\Delta$. We therefore investigate how $x_N$ compares with $x_\Delta$. We have $x_N > x_\Delta$ when

$$-d + \frac{1}{3}Rd(1+g+f_1) > 0, \tag{7.45}$$

which when using (7.32) and (7.34) gives the condition as

$$-1 + R\left[\frac{2(d-1)\sqrt{1-d^2}-1}{2[1+2\sqrt{1-d^2}]^2}+\frac{1}{2}\right] > 0. \tag{7.46}$$

Multiplying by a factor of $2[1+2\sqrt{1-d^2}]^2$ and collecting terms the condition

$x_N > x_\Delta$ can be reduced to

$$z_{10} := \sqrt{1-d^2}\,[R(1+d)-4] - 2(R-2)d^2 + 2R - 5 > 0. \tag{7.47}$$

Using Figure 7.8 we observe that $z_{10} < 0$ for $R \leq 2$ meaning that $x_\Delta > x_N$ when there are 2 or fewer rings. In the instance of $R \geq 3$ there is a cut-off value where $z_{10}$ changes sign indicating that there is a region where $x_N > x_\Delta$ and a region where $x_\Delta > x_N$. When $R = 3$, (7.47) is equivalent to (7.37) and therefore this cut off is $d_1^*$. When $R = 4$, we solve for the roots of $\sqrt{1-d^2}[4d] - 4d^2 + 3$ which gives

$$d_4^* \approx 0.9776.$$

We know from Figure 7.8 that this cut off value, $d^*$, increases as $R$ increases so we conclude that for $R \geq 4$ there exists $d^* > d_1^*$ such that passing this threshold value $x_\Delta > x_N$.

A summary of findings can be found below in Table 7.1.

| R | PageRank | |
|---|---|---|
| 1 | $x_1 > x_\Delta > x_L > x_\delta > x_N$ | $\forall d \in (0, d_1^*)$ |
| | $x_\Delta > x_1 > x_L > x_\delta > x_N$ | $\forall d \in (d_1^*, 1)$ |
| 2 | $x_1 > x_\Delta > x_L > x_\delta > x_N$ | $\forall d \in (0, d_2^*)$ |
| | $x_1 > x_\Delta > x_L > x_N > x_\delta$ | $\forall d \in (d_2^*, d_3^*)$ |
| | $x_1 > x_\Delta > x_N > x_L > x_\delta$ | $\forall d \in (d_3^*, d_1^*)$ |
| | $x_\Delta > x_1 > x_N > x_L > x_\delta$ | $\forall d \in (d_1^*, 1)$ |
| 3 | $x_1 > x_N > x_\Delta > x_L > x_\delta$ | $\forall d \in (0, d_1^*)$ |
| | $x_\Delta > x_N > x_1 > x_L > x_\delta$ | $\forall d \in (d_1^*, 1)$ |
| $\geq 4$ | $x_N > x_1 > x_\Delta > x_L > x_\delta$ | $\forall d \in (0, d_1^*)$ |
| | $x_N > x_\Delta > x_1 > x_L > x_\delta$ | $\forall d \in (d_1^*, d^*)$ |
| | $x_\Delta > x_N > x_1 > x_L > x_\delta$ | $\forall d \in (d^*, 1)$ |

Table 7.1: *R rings with directed shortcut and directed hub*: PageRank centrality for $x_1, x_\Delta, x_L, x_\delta$ and $x_N$ from the heuristic analysis. In the case $R = 4$: $d^* = d_4^*$.

Figure 7.9 shows the ratio of $x_1$, $x_\Delta$, $x_L$ and $x_\delta$ to $x_N$ as the free parameter $d$ is varied across its domain for $1 \leq R \leq 4$. To do this, we used $m = 1{,}000$ with $L = 501$ and solved the PageRank system (3.5) directly. By examining each of these figures we can see that the results agree with those from our heuristic analysis as presented in Table 7.1. In particular, Figure 7.9a shows a change in ranking between nodes 1 and $\Delta$ on passing the threshold $d_1^*$. Figure 7.9b shows a change in ranking between nodes $\delta$ and $N$; $L$ and $N$; 1 and $\Delta$ on passing the thresholds $d_2^*$, $d_3^*$ and $d_1^*$ respecitively. In addition, Figure 7.9c shows the ranking change between nodes 1, $N$ and $\Delta$ on passing the threshold $d_1^*$. In addition to Figure 7.9a – 7.9c, Figure 7.9d also shows the change in ranking between nodes 1 and $\Delta$ on passing $d_1^*$ – this change occurs independent on the number of rings in the network. We note here the non-monotonic behaviour of these ratios and whilst it is not particularly clear due to the scale, the ratio $x_1/x_N$ in Figure 7.9c becomes negative on passing $d_1^*$ where it then reaches a turning point before approaching 1 as $d \to 1$.

To completely characterise the PageRank rankings on this network for all choices of parameter $d$ we now compare node $N$ to those nodes in the ring close to the spike nodes 1 and $L$. This is valid for $R = 2$ where $d_3^* < d < 1$ and $R = 3$ where $d_1^* < d < 1$ where the general node $k$ is ranked somewhere equivalent to a node near either nodes 1 and/or $L$. We therefore consider where $x_k > x_N$ which gives

$$1 + gu_1^{p(k)} + g'u_1^{p_L(k)} > 1 - d + \frac{1}{3}Rd(1 + g + f_1).$$

Using (7.32), (7.33) and (7.34) this gives

$$\begin{aligned} \frac{2(d-1)\sqrt{1-d^2}-1}{[1+2\sqrt{1-d^2}]^2}u_1^{p(k)} \quad &- \quad \frac{1}{1+2\sqrt{1-d^2}}u_1^{p_L(k)} \\ &> \quad -d + Rd\left[\frac{1}{2} + \frac{2(d-1)\sqrt{1-d^2}-1}{2[1+2\sqrt{1-d^2}]^2}\right], \end{aligned}$$

which when multiplied by a factor of $[1 + 2\sqrt{1-d^2}]^2$ rearranges to

$$\begin{aligned} &\left[2(d-1)\sqrt{1-d^2}-1\right]u_1^{p(k)} - \left[1+2\sqrt{1-d^2}\right]u_1^{p_L(k)} \\ &> \left[(R-4)d + Rd^2\right]\sqrt{1-d^2} - 2(R-2)d^3 + (2R-5)d. \end{aligned}$$

We first consider nodes close to node 1 and note that when we are close to node

1, $u_1^{p_L(k)} = 0$. The condition (7.48) reduces to $p(k) > p(k)_{\min}$ where

$$p(k)_{\min} = \begin{cases} \dfrac{\log\left[2d(1-d)\sqrt{1-d^2}+d\right] - \log\left[2(1-d)\sqrt{1-d^2}+1\right]}{\log\left(1-\sqrt{1-d^2}\right)-\log d}, & \text{(7.48a)} \\ \dfrac{\log\left[d(1-3d)\sqrt{1-d^2}+2d^3-d\right] - \log\left[2(1-d)\sqrt{1-d^2}+1\right]}{\log\left(1-\sqrt{1-d^2}\right)-\log d}. & \text{(7.48b)} \end{cases}$$

Equation (7.48a) is valid for $R = 2$ and $d_3^* < d < 1$ whilst (7.48b) is valid for $R = 3$ and $d_1^* < d < 1$.

Moving to now consider nodes close to node $L$ we note that $u_1^{p(k)} = 0$ and the condition (7.48) reduces to $p(k) > p_L(k)_{\min}$ where

$$p_L(k)_{\min} = \begin{cases} \dfrac{\log\left[2d(1-d)\sqrt{1-d^2}+d\right] - \log\left[1+2\sqrt{1-d^2}\right]}{\log\left(1-\sqrt{1-d^2}\right)-\log d}, & \text{(7.49a)} \\ \dfrac{\log\left[d(1-3d)\sqrt{1-d^2}+2d^3-d\right] - \log\left[1+2\sqrt{1-d^2}\right]}{\log\left(1-\sqrt{1-d^2}\right)-\log d}, & \text{(7.49b)} \end{cases}$$

(7.49a) is valid for $R = 2$ and $d_3^* < d < 1$ whilst (7.49b) is valid for $R = 3$ and $d_1^* < d < 1$.

Figure 7.10 illustrates the functions (7.48a) – (7.49b).

The case of two rings in the network is illustrated in Figures 7.10a and 7.10b. This shows that $p_(k)_{\min} < 1$ and therefore all nodes close to 1 are deemed more important than the central hub node $N$ for values of the parameter $d$. Interestingly, $p_L(k)_{\min} < 1$ for $d$ chosen very close to $d_3^*$ before increasing and reaching an asymptote as $d \to 1$. It then follows that there is a choice of $d$ just marginally bigger than $d_3^*$ such that all nodes close to node $L$ will be more important than node $N$. However, as $d$ is increased so does $x_N$ at a rate faster than $x_k$ and therefore there are values of $d$ really close to 1 such that only nodes with $p_L(k) > P$ will be more important than then central hub node $N$.

The case of three rings in the network is illustrated in Figures 7.10c and 7.10d. Both of these functions have the same characteristic shape with both approaching an asymptote as $d \to d_1^*$.

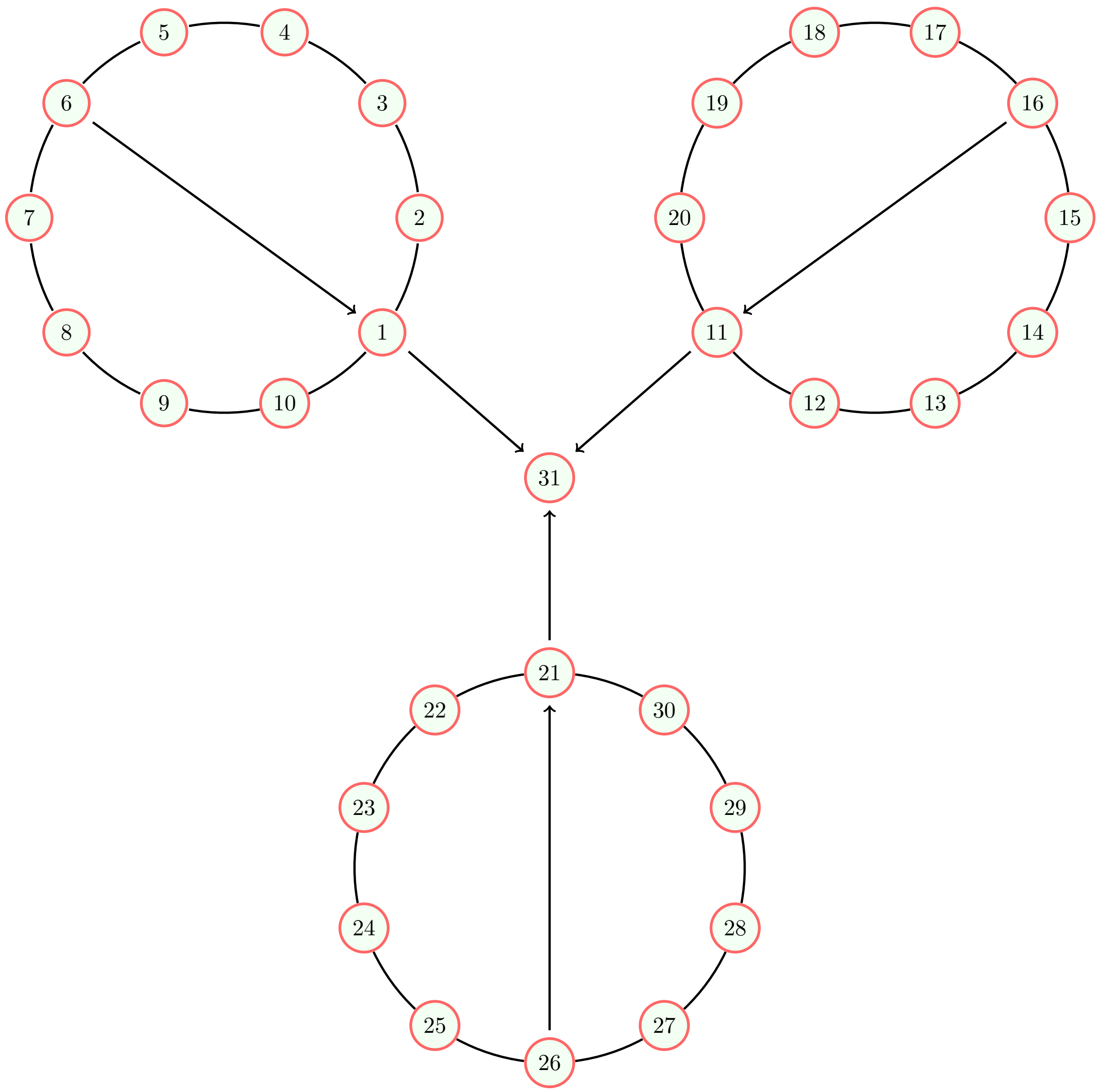
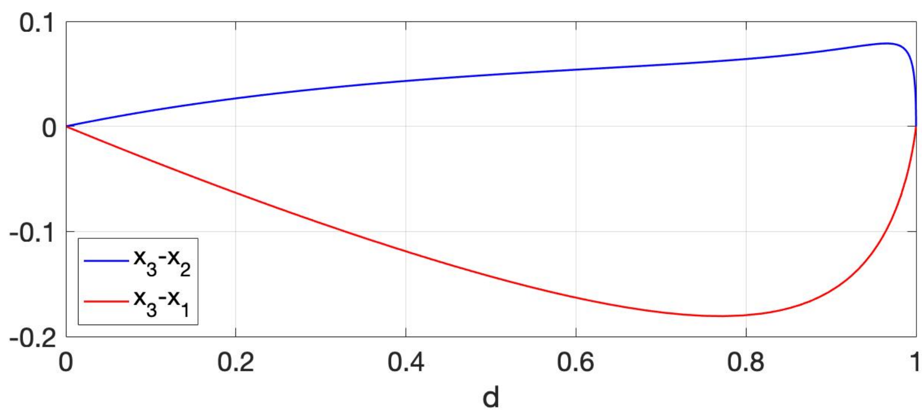



Figure 7.3: *R rings with directed shortcut and directed hub*: Illustration with directed shortcut and directed hub ($R = 3$, $m = 10$ and $L = 6$).



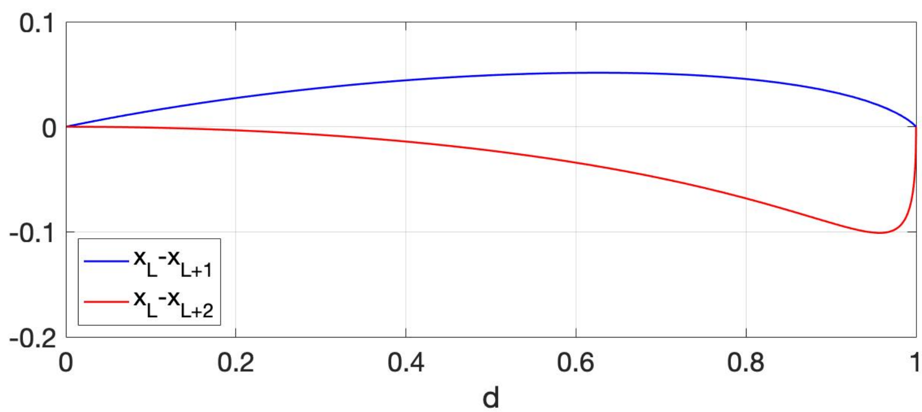
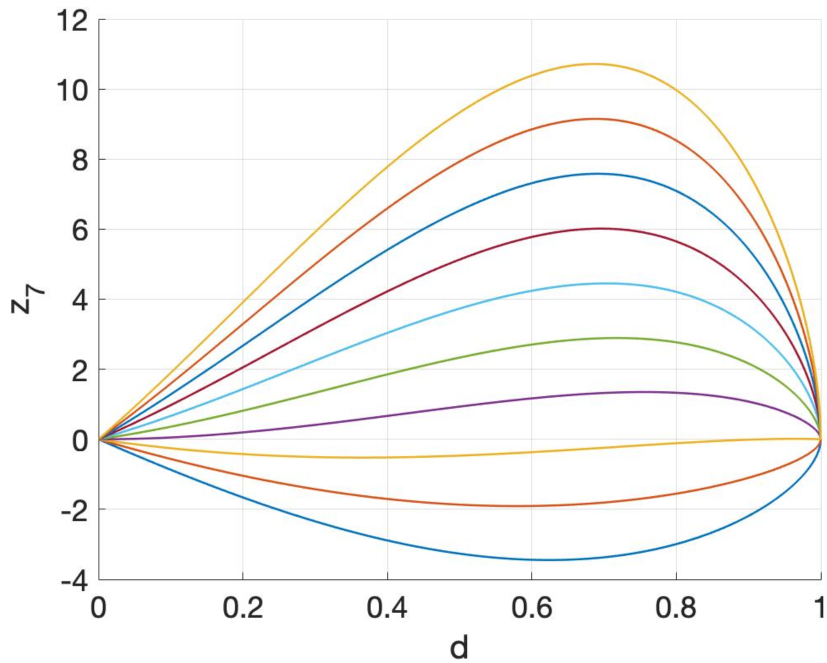
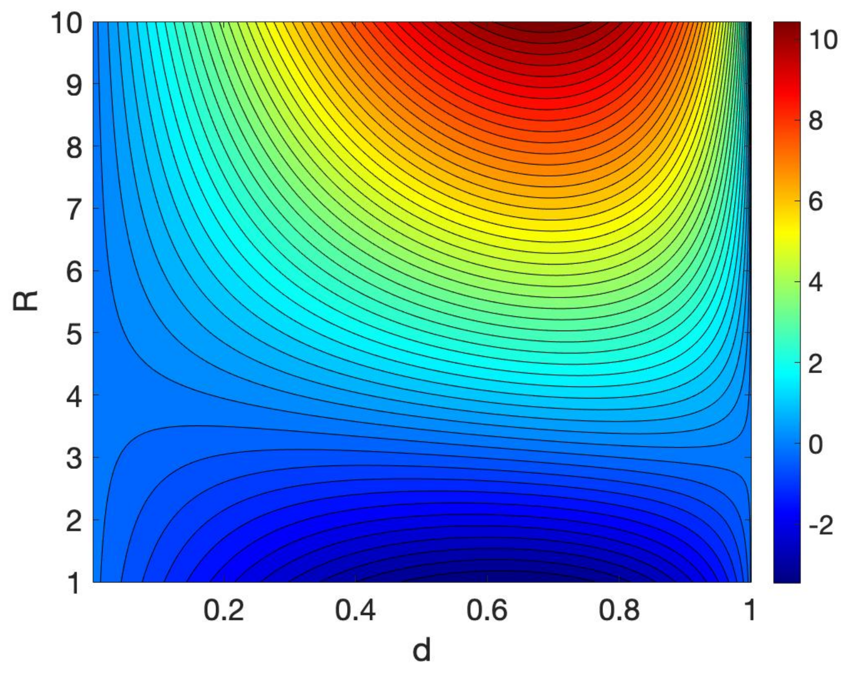




Figure 7.4: *R rings with directed shortcut and directed hub*: Difference in PageRank centrality for nodes that neighbour nodes 1 and $L$

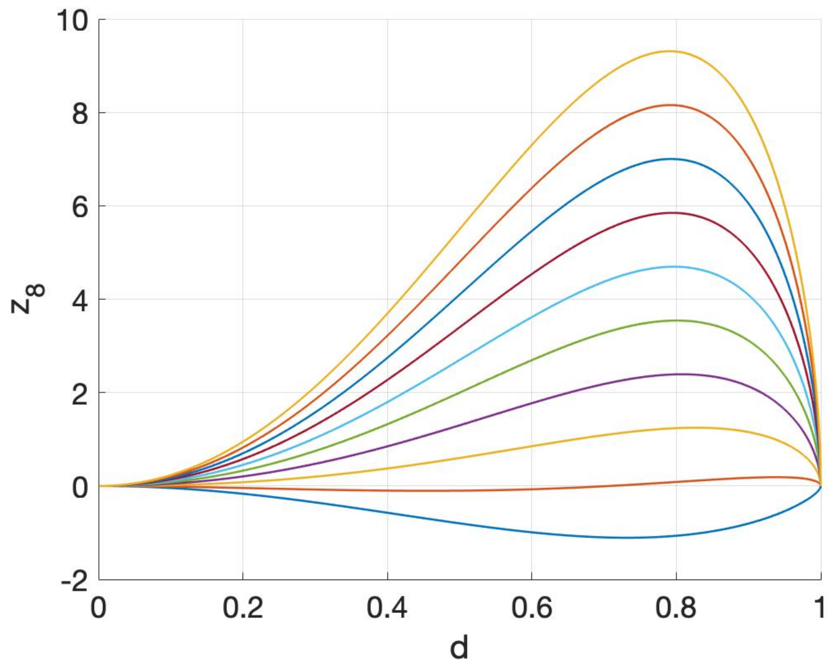



(a) Fixed $R$ ($R$=seq$(1, 10)$). $R = 1$ – bottom curve.

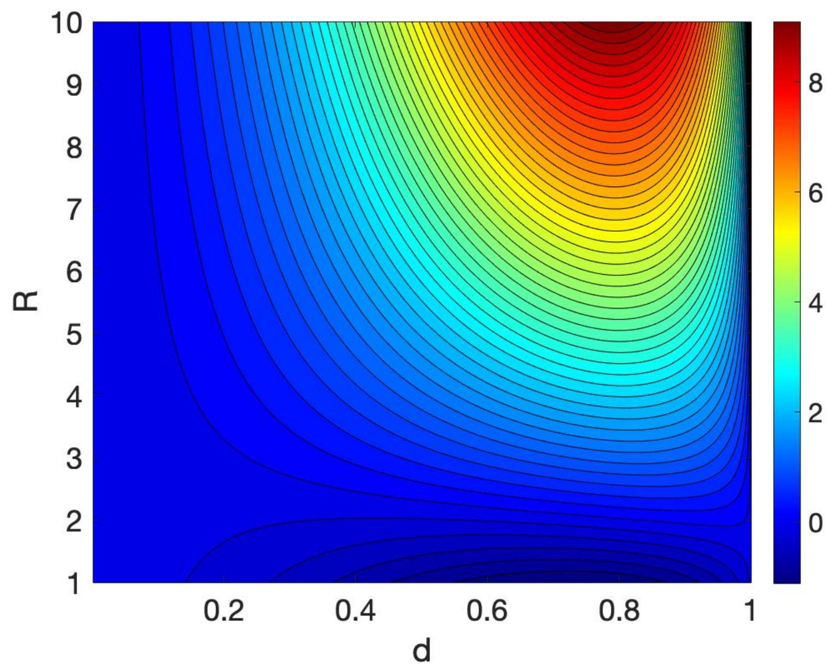



(b) Contour plot of $z_7$.

Figure 7.5: *R rings with directed shortcut and directed hub*: Relationship between number of rings $R$, PageRank parameter $d$ and $z_7$ in (7.40).

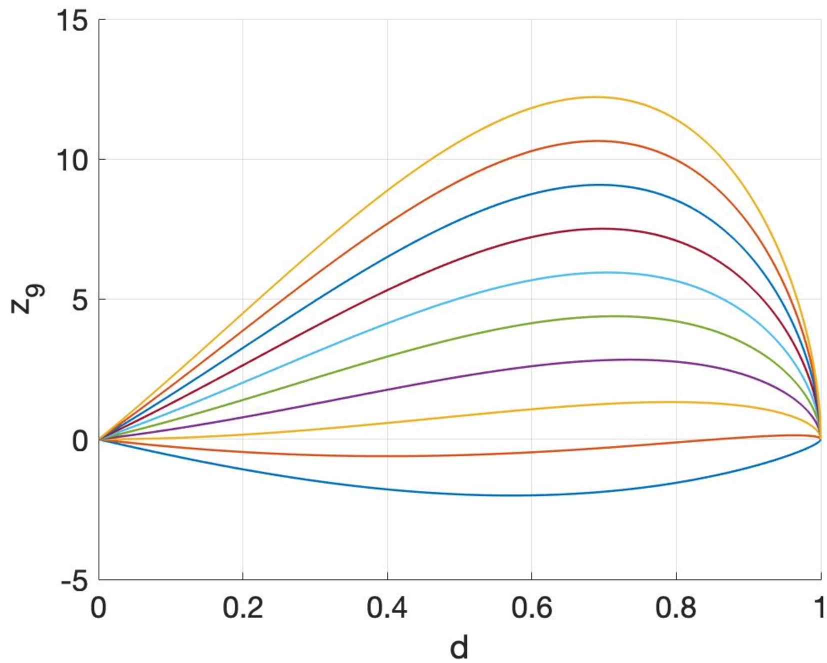



(a) Fixed $R$ ($R$=seq$(1, 10)$). $R = 1$ – bottom curve.

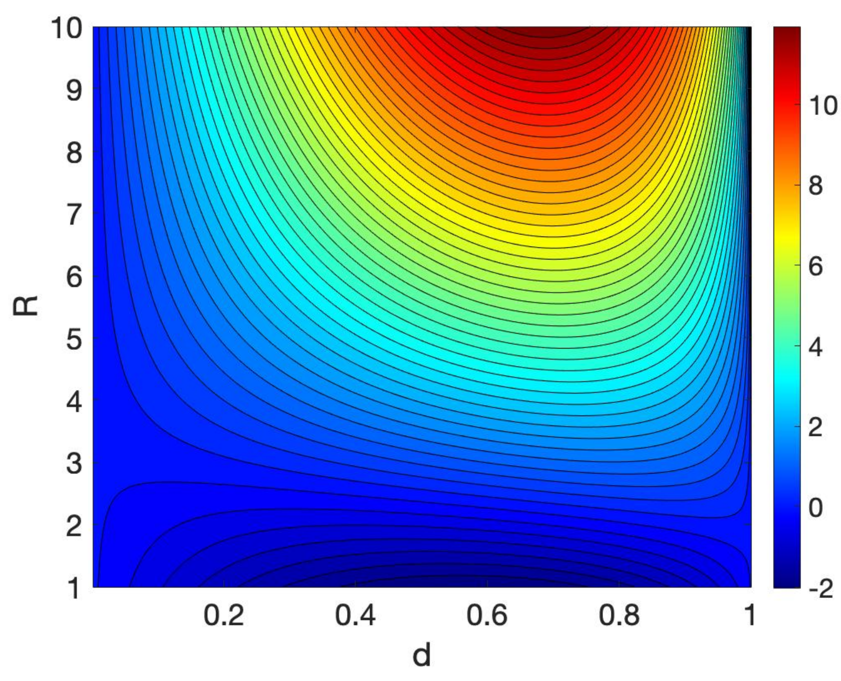



(b) Contour plot of $z_8$.

Figure 7.6: *R rings with directed shortcut and directed hub*: Relationship between number of rings $R$, PageRank parameter $d$ and $z_8$ in (7.42).

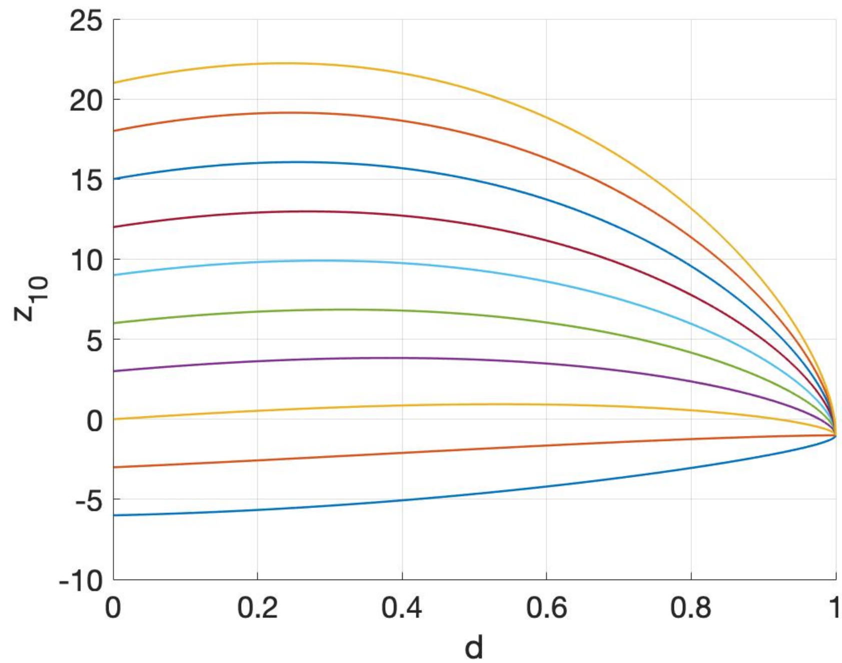



(a) Fixed $R$ ($R$=seq$(1, 10)$). $R = 1$ – bottom curve.

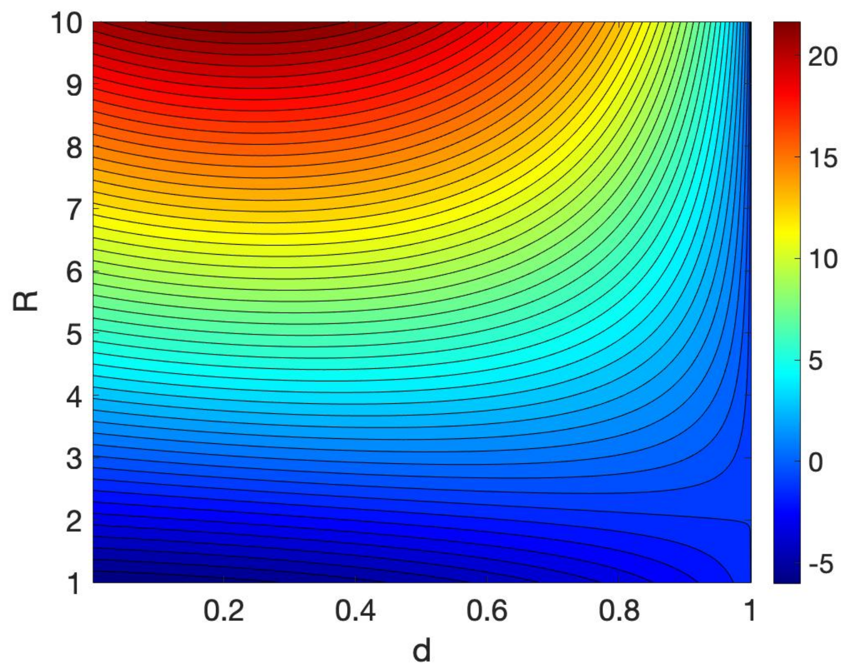



(b) Contour plot of $z_9$.

Figure 7.7: *R rings with directed shortcut and directed hub*: Relationship between number of rings $R$, PageRank parameter $d$ and $z_9$ in (7.44).

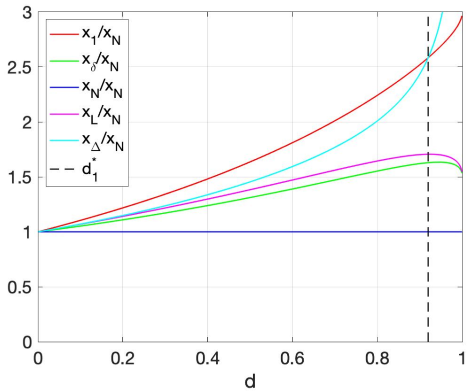



(a) Fixed $R$ ($R$=seq$(1, 10)$). $R = 1$ – bottom curve.

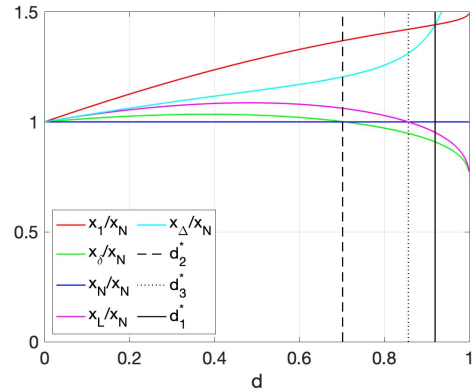



(b) Contour plot of $z_{10}$.

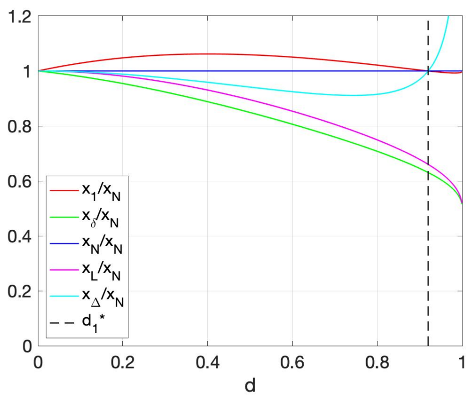
Figure 7.8: *R rings with directed shortcut and directed hub*: Relationship between number of rings $R$, PageRank parameter $d$ and $z_{10}$ in (7.47).



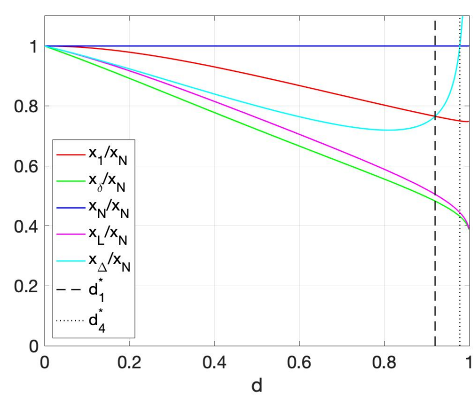

(a) $R = 1$




(b) $R = 2$

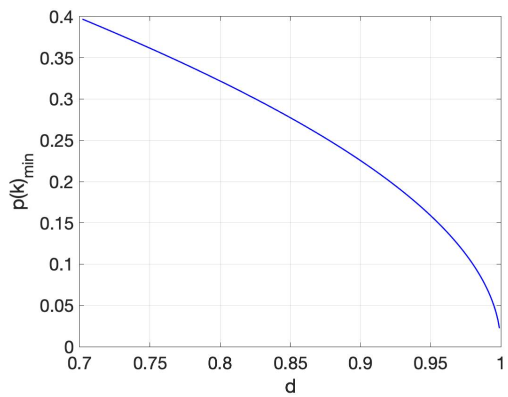



(c) $R = 3$

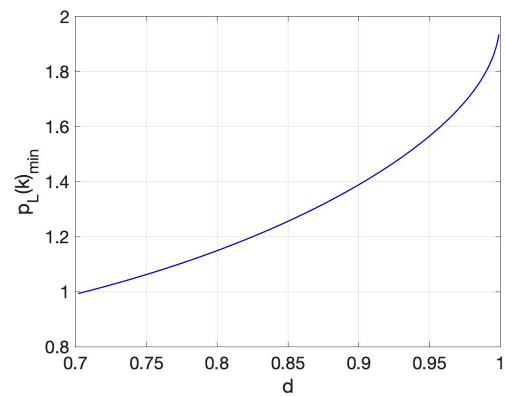



(d) $R = 4$

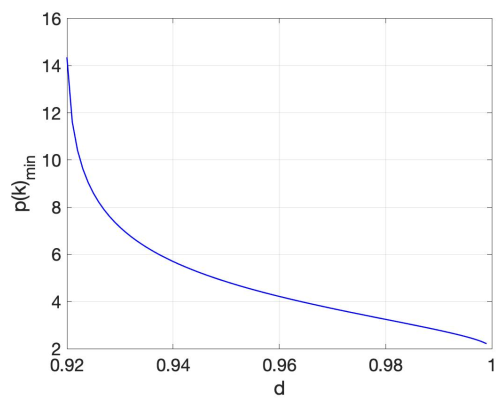
Figure 7.9: *R rings with directed shortcut and directed hub*: Ratio of $x_1$, $x_\delta$ and $x_L$ to $x_N$ for PageRank centrality. We used $m = 1,000$ so $N = 1000R + 1$ with $L = 501$.



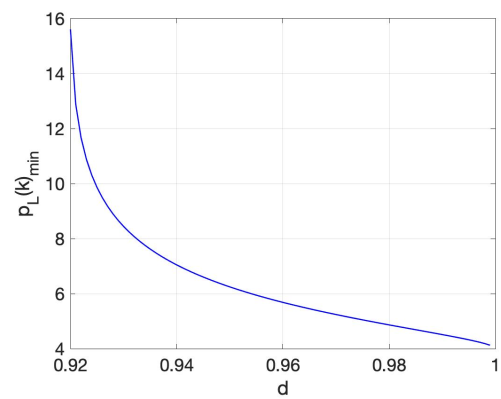

(a) $p(k)_{\min}$ for $R = 2$ and $d_3^* < d < 1$




(b) $p_L(k)_{\min}$ for $R = 2$ and $d_3^* < d < 1$




(c) $p(k)_{\min}$ for $R = 3$ and $d_1^* < d < 1$




(d) $p_L(k)_{\min}$ for $R = 3$ and $d_1^* < d < 1$

Figure 7.10: *R rings with directed shortcut and directed hub*: Illustrating the minimum periodic distance for which $x_k > x_N$

## 7.3 Summary of key results

In this chapter we considered PageRank centrality which is different from that of Katz and eigenvector centrality in that it is concerned with incoming information as opposed to outgoing information. We have shown the PageRank centrality vector form of solution on a nearest neighbour periodic ring with a single directed shortcut, illustrating a positive spike at the shortcut receiving node and an inverted spike and shift at the shortcut originating node. What was perhaps most surprising was the shift at the originating shortcut node which gives rise to an unexpected non-monotonicity in the nodal rankings.

We then applied this result to a more complicated network similar to that presented in Chapter 6 where we have $R$ nearest neighbour rings with a single directed shortcut connected to a central hub node. In this network configuration the key nodes are 1 (the target of the shortcut and source of the edge that connects to the central hub), $\Delta$ (a baseline node where no spike has any influence), $L$ (the source of the shortcut), $\delta$ (the minimum in the ring – one of $L$'s neighbours) and $N$ (the central hub).

Our analysis allowed us to show analytically the cut off values where nodal rankings changed in the network, hence illustrating the influence of the PageRank parameter $d$. In particular we found:

- Node 1 changes to being the tip of an inverted spike rather than a positive spike on passing the threshold value $d_1^*$ – irrespective of $R$. This means that we have a change in rank of nodes 1 and $\Delta$.
- When $R = 2$ there is a cut off value $d_2^*$ where the rank of nodes $\delta$ and $N$ change and there is a cut off value $d_3^*$ where the rank of nodes $L$ and $N$ change.

Furthermore, in the regions where node $N$ is neither ranked first nor last in the overall network our analysis shows where node $N$ ranks relative to the other nodes in the ring.

# Chapter 8

# Discussion

In this work, we have addressed issues in matrix analysis concerning modified rings. In particular, we have extended the approximation technique from [37] in order to give exact solutions, and applied these ideas to spectral and linear system problems arising in network science, where the matrix modifications correspond to shortcuts in the classic Watts-Strogatz model.

In the case of a nearest neighbour periodic ring plus single directed shortcut, components of the Katz, eigenvector and PageRank centrality vector have the form

$$\begin{aligned} x_i^{Katz} &= b + h_1 t_1^{p(i)} + h_1 t_2^{p(i)-N}, \\ x_i^{EV} &= s_1^{p(i)} + s_2^{p(i)-N}, \\ x_i^{PgR} &= 1 + g_1 u_1^{p(i)} + g_1 u_2^{p(i)-N} + g_1' u_1^{p_L(i)} + g_1' u_2^{p_L(i)-N} + \epsilon f, \end{aligned}$$

respectively where $\epsilon = 1$ if $i = L$ and $\epsilon = 0$ otherwise.

Katz and eigenvector centrality exhibit a positive localised spike centred at the node where the shortcut originates. This geometrically decreases on moving away from the shortcut originating node. PageRank centrality on the other hand has a different feel and exhibits not one but two localised spikes – a positive spike centred at the node that receives the shortcut and a negative spike centred at the shortcut source node. What was not intuitively obvious to all three of these centrality measures was the presence of an asymptotically exponentially small increasing geometric component that accompanies every spike. Katz and eigenvector centrality have a single geometrically increasing component moving away from the shortcut source node. PageRank centrality has two geometrically increasing components – one centered at each of the nodes involved in the shortcut.

Katz and PageRank centrality have a free parameter $\alpha$ and $d$ respectively. We note that the parameters $\alpha$ and $d$, which influence the geometric growth and decay rates, are independent of the network size. By contrast, for eigenvector centrality the geometric decay and growth rates are dependent upon the network size. The smaller the parameter the more localised the spike and the greater the parameter the more non-localised the spike will be. The parameter therefore affects the benefits that each node gets from a shortcut located a distance away in terms of the centrality score but does not influence the node rankings. PageRank centrality reveals an unexpected non-monotonicity in the rankings due to a shift at the shortcut originating node for which it was shown that this shift is enough to make the originating node more important than its neighbours but not more than its neighbours neighbour.

These three centrality measures all have a difference in their baseline value, that is, their value where the effect of a localised spike is negligible. Far from the originating node, Katz centrality tends to a fixed non-zero $b$ which depends on the parameter $\alpha$, eigenvector centrality tends to zero and PageRank centrality tends to 1 independently of the parameter $d$.

We extended the Katz centrality measure to the class of networks defined by (a) a nearest neighbour periodic ring with $M$ shortcuts and (b) a $2n$-nearest neighbour periodic ring with directed shortcuts. This revealed that in the case of multiple shortcuts the Katz vector is a sum of $M$ spikes each centred at the shortcut originating nodes plus a fixed non-zero value – each spike decaying and increasing at the same rate. In the case of a a $2n$-nearest neighbour periodic ring with shortcut the Katz vector is a sum of $n$ geometric decreasing terms, $n$ geometric increasing terms and a fixed non-zero value.

These results give a complete understanding of the structure of the centrality vectors on this class. In particular, our analysis of more general networks where the node ranking is not predictable in advance provides the first analytically derived examples where a threshold exists for the Katz and PageRank parameter across which the node ranking changes.

This work can be applied to software verification, for example, if we have an algorithm that proposes to compute a centrality score for nodes in a network. How can one determine if the algorithm output is correct? By having nontrivial example networks where we know exact expressions for centrality vectors means the implementation of the algorithm can be tested and verified before turning to large scale real-world application areas.

Additionally, in many real-world application areas, such as protein interactions, there is a well known problem of how to label nodes in a way that reveals structure that may not be apparent when nodes are labelled arbitrarily. These networks often result in adjacency matrices that can be arranged into particular forms such as those with entries dominated around the diagonal. In many of these areas data is collected experimentally and therefore there may be an issue with false positive results. Insight from our work may suggest that these false positive edges present in the adjacency matrix may result in localised spikes. Detecting such spikes may therefore aid in identification of these erroneous results.

Furthermore, these results give insight into the effect of adding shortcuts to a network. The addition of shortcuts across a network will often incur cost to implement and this work gives insight into where the benefit of such investment would be felt. For example, consider the Glasgow subway system – a nearest neighbour periodic ring network – and suppose that there were proposals to add new tracks between existing stations. Insight from this work would suggest this would result in a localised spike in importance of stations close to where this shortcut originates (and terminates if trains can flow in both directions). This allows investors to consider what they want to achieve and how best to achieve it.

There is clearly much potential for further work in this area. Whilst we have provided conjectures based on $M$-shortcuts across a nearest neighbour periodic ring and the $2n$-nearest neighbour periodic ring plus shortcut it would be nice to be able to definitively prove these.

In a $M$-shortcut ring questions such as

- how well separated should shortcuts be,
- is there a limit to the number of shortcuts that can be added before the geometrically decaying spike structure is lost,
- what are the precise asymptotics of the Perron-Frobenius eigenvalue,
- can we study a limit where the number of shortcuts $M$ and the number of nodes $N$ tend to infinity at differing rates,

could be addressed.

A more immediate issue in the $2n$-nearest neighbour ring plus shortcut would involve an understanding of the roots of the palindromic polynomial $r^n - \alpha(r^{2n} + \cdots + r^{n+1} + r^{n-1} + \cdots + 1)$ which can give rise to complex rates of geometric decay and growth such as in the case discussed at the end of Section 5.2 where $n = 3$ and $\alpha = 0.1 < 1/\rho(A)$. Whilst we hypothesise that as these complex rates appear in groups of four and counterbalance imaginary components we feel that this is an interesting direction for further investigation.

As these extensions were provided as a basis of justification for the more complex examples in Chapter 6, these ideas could be combined to consider the case of $M$-shortcuts across a $2n$-nearest neighbour periodic ring. As we have explored network extensions under the Katz centrality these could be equally be examined for other centrality measures such as eigenvector and PageRank.

Furthermore the questions addressed here could be posed on (a) higher dimensional lattices, such as nearest-neighbour connections on a torus, (b) hierarchical "network of network" structures, or (c) non-periodic lattices where boundary effects may be significant. In these cases, the underlying linear algebra challenges involve spectral analysis of Toeplitz matrices with various block structures and low rank perturbations, building on ideas in [13, 14].

Multiple shortcuts could be included, with various asymptotic scalings for their length and separation. Also, in the spirit of [50], random shortcuts could be analysed in a suitable probabilistic framework. Extensions could also be considered to a temporal network setting where the edges are time-dependent by extending the centrality measures in a manner described in [27, 47], or to the case of weighted networks.

# Bibliography


[1] W. Ahrens, *Zur relativen bewertung von turnierpartien*, Wiener Schachzeitung, 4 (1901), pp. 181–192.

[2] M. Aprahamian, D. J. Higham, and N. J. Higham, *Matching exponential-based and resolvent-based centrality measures*, Journal of Complex Networks, 4 (2016), pp. 157–176.

[3] S. Aral, *Social science: Poked to vote*, Nature, 489 (2012), pp. 212–214.

[4] A. Barbour and G. Reinert, *Small worlds*, Random Structures and Algorithms, 19 (2001), pp. 54–74.

[5] A. Bavelas, *A mathematical model for group structures*, Applied Anthropology, 7 (1948), pp. 16–30.

[6] ——, *Communication patterns in task-oriented groups*, Journal of the Acoustical Society of America, 22 (1950), pp. 725–730.

[7] A. Bavelas and D. Barrett, *An experimental approach to organizational communication*, American Management Association, (1951), pp. 57–62.

[8] M. Benzi and C. Klymko, *On the limiting behavior of parameter-dependent network centrality measures*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 686–706.

[9] N. Boccard, *The Landau legacy: from chess to PageRank*, (2019).

[10] P. Boldi, A. Luongo, and S. Vigna, *Rank monotonicity in centrality measures*, Network Science, 5 (2017), p. 529–550.

[11] F. Bonchi, P. Esfandiar, D. F. Gleich, C. Greif, and L. V. Lakshmanan, *Fast matrix computations for pairwise and columnwise commute times and Katz scores*, Internet Mathematics, 8 (2012), pp. 73–112.

[12] A. Böttcher, M. Embree, and M. Lindner, *Spectral approximation of banded Laurent matrices with localized random perturbations*, Integral Equations and Operator Theory, 42 (2002), pp. 142–165.

[13] A. Böttcher, M. Embree, and V. I. Sokolov, *On large Toeplitz band matrices with an uncertain block*, Linear Algebra and its Applications, 366 (2003), pp. 87–97.

[14] ——, *The spectra of large Toeplitz band matrices with a randomly perturbed entry*, Math. Comp., 72 (2003), pp. 1329–1348.

[15] R. L. Burgess, *Communication networks and behavioral consequences*, Human Relations, 22(2) (1969), pp. 137–159.

[16] M. Catral, M. Neumann, and J. Xu, *Matrix analysis of a Markov chain small-world model*, Linear Algebra and its Applications, 409 (2005), pp. 126–146.

[17] T. P. Chartier, E. Kreutzer, A. N. Langville, and K. E. Pedings, *Sensitivity and stability of ranking vectors*, SIAM J. Scientific Computing, 33 (2011), pp. 1077–1102.

[18] B. S. Cohn and M. Marriott, *Networks and centres of integration in Indian civilization*, Journal of Social Research, 1 (1958), pp. 1–9.

[19] J. A. Czepiel, *Word-of-mouth processes in the diffusion of a major technological innovation*, Journal of Marketing Research, 11(2) (1974), pp. 172–180.

[20] P. Erdös and A. Rényi, *On random graphs*, Publ. Math. Debrecen, 6 (1959), pp. 290–297.

[21] E. Estrada, *The Structure of Complex Networks*, Oxford University Press, Oxford, 2011.

[22] L. Euler, *Solutio problematis ad geometriam situs pertinentis*, Commentarii Academiae Scientiarum Imperialis Petropolitanae, 8 (1736), pp. 128–140.

[23] L. C. Freeman, *Centrality in social networks conceptual clarification*, Social Networks, 1 (1978), pp. 215–239.

[24] E. N. Gilbert, *Random graphs*, Ann. Math. Statist., 30 (1959), pp. 1141–1144.

[25] D. F. Gleich, *PageRank beyond the web*, SIAM Review, 57 (2015), pp. 321–363.

[26] P. Grindrod, *Mathematical Underpinnings of Analytics*, Oxford University Press, Oxford, 2014.

[27] P. Grindrod and D. J. Higham, *A dynamical systems view of network centrality*, Proc. R. Soc. A, 470 (2014), p. 20130835.

[28] D. J. Higham, *A matrix perturbation view of the small world phenomenon*, SIAM Journal on Matrix Analysis and Applications, 25 (2003), pp. 429–444.

[29] D. J. Higham, *Greedy pathlengths and small world graphs*, Linear Algebra and Its Applications, 416 (2006), pp. 745–758.

[30] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 2nd ed., 2013.

[31] A. Jeffrey, *Complex analysis and applications*, Chapman and Hall/CRC, 2005.

[32] L. Katz, *A new index derived from sociometric data analysis*, Psychometrika, 18 (1953), pp. 39–43.

[33] E. Landau, *Zur relativen wertbemessung der turnierresultate*, Deutsches Wochenschach, 11 (1895), pp. 366–369.

[34] A. N. Langville and C. D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton, NJ, USA, 2006.

[35] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, and T. Jebara, *Computational social science*, Science, 323 (2009), pp. 721–723.

[36] H. J. Leavitt, *Some effects of certain communication patterns on group performance*, The Journal of Abnormal and Social Psychology, 36(1) (1951), pp. 38–50.

[37] X. Liu, G. Strang, and S. Ott, *Localized eigenvectors from widely spaced matrix modifications*, SIAM J. Discrete Math., 16 (2003), pp. 479–498.

[38] Z. Lu, B. Savas, W. Tang, and I. Dhillon, *Supervised link prediction using multiple sources*, in Data Mining (ICDM), 2010 IEEE 10th International Conference on, Dec. 2010, pp. 923 –928.

[39] S. Milgram, *The small world problem*, Psychology Today, 2 (1967), pp. 60–67.

[40] R. Monasson, *Diffusion, localization and dispersion relations on "small-world" lattices*, Eur. Phys. J. B, 12 (1999), pp. 555–567.

[41] M. E. J. Newman, *Networks: an Introduction*, Oxford Univerity Press, Oxford, 2010.

[42] M. E. J. Newman, C. Moore, and D. J. Watts, *Mean-field solution of the small-world network model*, Phy. Rev. Lett., 84 (2000), pp. 3201–3204.

[43] M. E. J. Newman and D. J. Watts, *Renormalization group analysis of the small-world network model*, Physics Letters A, 263 (99), pp. 341–346.

[44] L. Page, S. Brin, R. Motwani, and T. Winograd, *The PageRank citation ranking: Bringing order to the web*, Technical Report, Stanford University, 1998.

[45] M. Paton, K. Akartunali, and D. J. Higham, *Centrality analysis for modified lattices*, SIAM J. Matrix Anal. Appl., 38(3) (2017), pp. 1055–1073.

[46] H. Sun, Y. Wang, and H. X. Zhang, *Polynomials with palindromic and unimodal coefficients*, Acta Mathematica Sinica, English Series, 31 (2015), pp. 565–575.

[47] D. Taylor, S. A. Myers, A. Clauset, M. A. Porter, and P. J. Mucha, *Eigenvector-based centrality measures for temporal networks*, Multiscale Modeling and Simulation, 15 (2017), pp. 537–574.

[48] S. Vigna, *Spectral ranking*, Network Science, 4 (2009).

[49] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, 1994.

[50] D. J. Watts and S. H. Strogatz, *Collective dynamics of 'small-world' networks*, Nature, 393 (1998), pp. 440–442.

[51] J. Xu, *Markov chain small-world model with asymmetric transition probabilities*, Electronic Journal of Linear Algebra, 17 (2008), pp. 1–22.