7-31-2016

Communication Network Among Campus Sustainability Influencers

Alexandra C. Freling
Loyola Marymount University, afreling@lion.lmu.edu

Communication Network Among Campus Sustainability Influencers

Coco Freling

Honors Thesis

in Partial Fulfillment of the Requirements for the

LMU Honors Program

Faculty mentor: Dr. Dorothea Herreiner

Economics

July 2016

Acknowledgements

I owe an immeasurable amount of gratitude to Dr. Dorothea Herreiner, my advisor on this project. I tend to present over the top ideas and ambitious hypotheses that take too much time and that are well beyond my breadth of knowledge. However, without question, you supported me as I attempted to link various aspects of my life and interests into one study. You challenged me to go beyond simple observation to learn and really understand every supporting theory, method, and idea from which my research extends. You have guided me through an area of study that I otherwise would not have had the opportunity to learn, and I am so glad I did. My interest in network analysis has only grown from this experience.

I also want to thank other supporting faculty and staff who provide guidance over the course of this project. Dr. Healy, perhaps unknowingly, your Advanced Econometrics class helped me throughout the course of my research. With the skills you taught me, I was able to not only questions the data in the articles that I read, but actually run tests and validate results myself. April and Viktoria, you both helped give light to this project for so many different reasons. Without you two, not only would this research not exist, but I would not have had the confidence to stand up for change.

Bryce and Richard, during my times of total disarray with trying to decipher and recode very unfamiliar MatLab and LaTeX code, you both came to my rescue. Truly truly, I owe so much of my success in the paper to you both and the generous amount of time you both dedicated to helping me. Mom and Dad, thank you for the endless hours of support and paper editing especially as I hit a time crunch.

To all the staff, faculty, and students who participated in my research thank you. Your insight has opened up a new window for discussion and possibilities. All in all, I could not have accomplished this work without the encouragement of these tremendous people.

Abstract

Systems of all types require efficient communication between its parts and units in order to be successful and effective. It is thus important to understand a systems units in order to better advance its operations. In this study, we look at Loyola Marymount University (LMU) as a systematic organization in regards to the universitys execution of its environmental sustainability endeavors. This approach allows for the identification of the path by which important environmental sustainability information is communicated, is learned, and is acted upon at LMU. Through various network centrality measurements, I will develop a visual representation of the communication network between individuals on LMU's campus who have an interest and play a role in the development and advancement of environmental sustainability practices and policies on campus. Moreover, an analytical understanding of this network of information transference will provide insight into the decision-making, implementation, and management that affects the efforts to reduce LMU's campus carbon footprint.

"If you want to bring a fundamental change in people's belief and behavior...you need to create a community around them, where those new beliefs can be practiced and expressed and nurtured." Malcolm Gladwell,

The Tipping Point:

How Little Things Can Make a Big Difference

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Setting the stage

Network analysis and the impact of information "spreading" are common scientific research fields. Their applications, however, are ever changing, leading to studies about new methods of understanding networks, connectedness, and information contagion. Theories and classic topics are being revisited, questioned, and adapted by those who study the fields. Globalization, technology, and environmental change offer us developing case studies to better explore and understand the notion of connectedness.

In high school, I researched the influence of early advertising on the effects of cigarette consumption. Inspired by Malcolm Gladwell's insights about how trends spread via word of mouth in his book *The Tipping Point*, I decided to consider the influence of advertisements as a social system. My interest in this led me to pursue my degree in Marketing at Loyola Marymount University (LMU). As a result of my marketing research classes, I learned that in order to fully understand the implications and effects of the advertising it is necessary to think about the impacts as a network, as a trackable structure of contagion and influence.

Outside of classes, I work at both a market research firm and collectively at LMU's Center for Urban Resilience (CURes) and Green LMU. In both roles, I was asked to identify and question the ways in which people respond to the transmission of information. My position with CURes and Green LMU fostered a specific interest about how environmental sustainability trends and programs stick (or don't stick) socially. The integration of my two roles has kept me questioning continuously how environmental sustainability trends and programs stick. Beyond the confines of LMU, Los Angeles, as a whole, is an environmentally conscious city. I see Save the Drop advertisements all around the city. I see community gardens in almost every neighborhood and monthly beach clean up events. My hometown of Dallas, Texas, on the other hand, has a very different culture with regards to eco-friendliness. I go home and see fewer recycling bins, have little access to local farmers' markets, and follow heavy-duty pickup trucks at every intersection. Naturally, I have wondered why the important sustainability trends "stick" in L.A. but not Dallas.

I believe that people in Dallas are just as environmentally conscious as those in L.A. but the thoughts, knowledge, support, and affinity seem to have a harder time translating into action. The gap between supporting something and becoming an active advocate is an important research area. For instance, in the world of advertising, researchers seek to understand what influences a conversion from intent to purchase to actual transaction. For now, Gladwell suggests the best we can do is to identify network influencers who might inspire a discernable tip in dissemination of information.

Through a series of coincidentally timed events between my work at the research firm, CURes, and my Econ 530 class at LMU, I became curious about the generation and cohesion of environmental sustainability at LMU, and about the spread of information regarding LMU's environmental sustainability. I approached Dr. Dorothea

Herreiner, my Econ 530 professor, about studying LMU's environmental sustainability operations through a communication network analysis for my senior thesis.

My personal goal in doing this research is twofold. One, I have spent my four years at LMU working in different Facilities Management positions around campus. Within the department, the biggest point of concern has been disconnectedness with other individuals, departments, and groups campus. After speaking to professors and staff members in various campus position, and reading testimonials in the Loyolan, I found this to be a recurring issue throughout the entire university. The siloing of groups and organizations is not only hurting the community culture on campus, but also inhibiting progress in many areas, especially sustainability, which requires a collaboration and team effort. Two, I am interested in learning about networks and network analysis, finding it applicable to many different areas of my life and areas of focus. So, for myself, I hope to gain a better understanding about LMU's organizational behavior and where communication or lack thereof impacts the collaborative effort and unified mentality at the university. I aim to learn how formal and informal information communications and behaviors on a university campus lead to discernable insight into overall operative organizational success. Should decision makers consider my findings, I hope that LMU might be able to use my research to enhance its efforts towards becoming the most environmentally sustainable university campus.

Chapter 2

Environmental Sustainability (ES) at LMU

"Anything else you're interested in is not going to happen if you can't breathe the air and drink the water. Don't sit this one out. Do something." Carl Sagan,

In 2007, LMU faculty, staff, and students came together to build the Environmental Stewardship and Sustainability Committee (E2SC). Their goal was to unify campus goals, objectives, involvement and commitment to environmental justice and sustainability to maintain and grow their national environmental sustainability[1]leadership [34].

[1]Throughout my study, I refer to "involvement" and "environmental sustainability" according to the following definitions:

Involvement - Participation in the advancement of environmental sustainability operations and practices on LMUs campus, and all activity that helps to reduce the campus carbon footprint.

Environmental Sustainability - This definition is definition taken from Herman Daly: 1. Renewable resources, the rate of harvest should not exceed the rate of regeneration; 2. Pollution, the rates of waste generation from projects should not exceed the assimilative capacity of the environment; and 3. Nonrenewable resources, the depletion of the nonrenewable resources should require comparable development of renewable substitute for that resource.

Daly, H. E. 1990a. Boundless bull. Gannett Center Journal 4(3):113118. Daly, H. E. 1990b. Toward some operational principles of sustainable development. Ecological Economics 2:16. Jan. 24. `http://www.thwink.org/sustain/glossary/EnvironmentalSustainability.htm`

Over the next seven years, with the support of this committee, LMU continued to seek out ways to develop, improve, and enhance their environmental sustainability efforts and programs. Almost every department on campus participated in ways most fitting to its area of activity, and they collaborated with other departments to build cross-campus initiatives and efficiencies. Now, in 2016, the E2SC meetings no longer convene. Still, many departments stand by their commitment towards incorporating environmental sustainability as a factor in all their decisions. Student Housing, for example, has the Student Housing Sustainable Purchasing Policy where they have pledged to purchase equipment and other materials and goods that are recyclable, reduce waste, promote energy efficiency, and careful waste segregation [34]. However, cross-department coalition appears to be weaker. This and other university disconnect may not be the reason for the discontinuation of the E2SE meetings, but I cannot help but wonder how this lost communication has affected (if at all) the viability of LMU's efforts for environmental sustainability innovation.

2.1 National Recognition and Standing Achievements

LMU boasts a number of recognizable environmental sustainability achievements. The university's automatic local weather sprinkler response system is an example:

- $\frac{3}{4}$ of the campus uses reclaimed water for irrigation
- 1st to have a campus-wide collegiate recycling program in California in 1993
- 1st California university to recycle 100% of its green waste
- In 2003, LMU won the federal Green Power Award in 2003 for its solar rooftop systems on University Hall and Von Der Ahe Building

- The campus has 5 LEED Certified buildings, with the Life Sciences Building earning a LEED Gold certification in 2015 and the William H. Hannon Library receiving a LEED Gold certification in 2011 ("LMU Facts," 2015).

In 2012, the university also received a Silver rating from AASHE Stars (The Association for the Advancement of Sustainability in Higher Education's Sustainability Tracking, Assessment and Rating System), which defines "sustainability" as "encompassing human and ecological health, social justice, secure livelihoods, and a better world for all generations" [20]. There is, however, so much more going on at LMU with regards to environmental sustainability beyond the banner advertisements. Personally, I believe that LMU should be most prideful not of these individual achievements, but rather their long-standing commitment and dedication to advancing university standards for environmental sustainability.

2.2 Brief History

In 2015, Pope Francis wrote *On Care For Our Common Home*, which speaks of the urgency for a collective effort to replenish, protect, and nurture the earth and its inhabitants [30]. His formal declaration about the validity of Global Warming generated a new wave of environmental sustainability consciousness. Ideas of sustainability and restoration, however, have long been embedded in the Jesuit tradition, and are, in fact, the reason that many Jesuits see environmental sustainability as a form of giving back. Bill Stonecypher, LMU's Manager of Facilities and Waste Management (better known as the Head Recycler) attributes his sustainability efforts and love for recycling to his Jesuit faith [44].

In 1992, LMU hired Stonecypher as the campus' Waste Recycling Coordinator. This was two years after the university built the first-ever university-wide recycling program in the country. Impressively, the development of LMU's recycling program came nine years before the General Congregration (GC) 35 where the supreme governing Jesuits would first discuss and endorse a commitment to ecological challenges in *A Broken World* [19]. Stonecypher believes that this anticipatory leadership came about because of the strong Jesuit principles fostered here at LMU.

As a requirement by the California Environmental Quality Act (CEQA), the City of Los Angeles with help from the firm Impact Science, competed a Draft Environmental Impact Report for LMU to accompany the 2010 Master Plan Project preceding their 20-year Development Plan [34]. This report details the environmental impact of new university plans and their respective solutions to mitigate any negative externalities of the plan. In order to build and affirm these solutions, LMU partners with and seeks guidance from outside organizations. Perhaps the most recognized of these organizations is the U.S. Green Building Council (USGBC) and their Leadership in Energy and Environmental Design [12]. In fact, LMU's new Life Science Center was built around the LEED Gold Certification guidelines, which specifies requisites on Location and Transportation accessibility of the building, Sustainable Site relationship with the ecosystem, Water Efficiency, Energy and Atmosphere, Materials and Resources, and Indoor Environmental Quality [12]. Additionally, our recycling center continues to win categories for the annual nationwide university Recyclmania competition [44].

2.3 Organizational Engagement

As is evident by the success of previous efforts and public statements, LMU, as a university, values environmentalism and conservationism. From my time at LMU, I have also noticed that this value permeates the university on the individual level. Many professors teach environmental sustainability practices and the involvement of environmental care in the everyday life. Others teach environmental sustainability through research or as an ethical humanitarian act. Departments arrange projects and programs such as garden workdays, campus-wide recycling, clothing swaps, and ES campus tours to further encourage education and good ES practices. Of recent, however, the practices advocated by LMU have become white noise. Concepts around the importance and impact of recycling and energy and water saving are now well known, especially in California. While I applaud LMU for continuing to promote and teach these concepts, LMU now lags behind other universities in green innovation and development, an area in which we were once leaders.

I believe that for LMU to truly stand as a leader in national ES efforts requires university-wide efforts and involvement. LMU faculty, staff, and student must collaborate and work in tandem with analogous objectives towards an overarching goal. This requires cross-departmental communication in order to ensure the alignment of projects, events, tasks, strategies, and efficiencies. For example, LMU Housing and the Sustainability Office are currently running a composting pilot study regarding the opportunity of developing a campus-wide composting program for students living in on-campus apartments [24]. This program requires support from Sodexo, LMU housing, ASLMU student leaders, student volunteers, Green LMU and the Sustainability Office, campus recycling, CURes, and Facilities Management. Before LMU moves beyond the pilot to offer a campus-wide composting program, the plan and pilot data and evaluations must first travel through a series of checks and balances for approval to ensure that the long-term benefits outweigh any inconveniences or costs.

To complete the approval, checks and balances process for any new ES decision or initiative, LMU departments clearly identify representatives and their expected roles in process. Like at any organization whether a formal business or a university, LMU details a set of expectations, tasks, and goals for different employment positions. Accordingly, the university has a handful of employees whose jobs specific focuses on ES programs and communication. For example: Trevor Wiseman coordinates

ES initiatives for LMU Housing, Bill Stonecypher represents and presents the efforts and initiatives of LMU Recylcing, Wassim Boustam works with LMU Food Services and speaks on our behalf with Sodexo, and Ian McKeown works as the campus' Sustainability Officer. If expectation for job roles and performance is clear, then active leadership by these individuals may reduce objectiveless attention.

2.4 Tribulations

ES development requires high short-term costs with the majority of the benefits visible in the long-term; and LMU may not realize these benefits directly, but instead only earn the prestige of being environmentally conscious. This perhaps reduces decision-makers' quickness and eagerness to implement ES programs. Instead, ES decisions may be placed on the backburner behind other more seemingly urgent issues. Even if a decision and project does pass, decision makers must identify signposts to demonstrate the progression of ES project along the way. Otherwise, the project is again subjected to attention flight and oversight, and may not produce a sustainable, long-term cultural change on campus.

Bekessy, Samson, and Clarkson (2007), researchers who have studied extensively the impact and value of environmental sustainability and non-binding environmental sustainability agreements at universities, also point out that because true benefits of ES programs are imperceptible at the micro level, universities might get rewarded for the idea without actually following through with the management of the project. Without signposts, there is no way to hold individuals accountable. Instead, individuals may only feel accountable if they are observed by minding person or persons. This, however, requires that more than one or a few individuals attend to the decision.

Like any multi-tiered operation, LMU's environmental sustainability efforts require the participation of the university leadership and different campus departments, the collaboration between leadership, departments and active students, and concise and persuasive messaging about common goals, objectives, and expectations before the university can successfully implement decisions and initiatives university-wide. Without clear collaboration with and the support of university management and administration, decision may never come to fruition. Recognition of the need for university-wide ES involvement is not enough, however, if those leading the university's ES development do not put forth steps to engage and educate other university members. Failure to engage the greater community, or at least matriculate the support of many campus departments, significantly reduces chance for successful program implementation and cultural change towards environmental sustainability. Part of this problem stems from generally poor communication around campus. I remember sitting in the CURes office one day, speaking with April Sandifer when she unexpected exclaimed, "Did you know LMU is hosting a feminist music and film festival this weekend?!" I had no idea, but even more interesting than my ignorance is that April learned about the event from LA Weekly, not LMU. We had to dig through LMU's website before finding any information about the event. When I walked around campus later that day, I only saw one display board announcing the event. One would presume that the school and hosts of the event would enthusiastically, loudly, and proudly promote the event for days before the event. Another time, previous to this experience, a pipe burst in the Lions Garden, which sits between CURes and GreenLMU. It took days before Facilities Management could find and fix the pipe because no one knew where to find the blue print for LMU's pipe system. Apparently, only one person knew where to find the information, and he had left LMU. It surprised me to learn that such important information, critical to the maintenance of the school was not known or at least easily accessible. These are just few of the many times I heard of and experienced poor communication at LMU. It is also clear to me that this problem is not confounded to the context of environmental sustainability, but rather, a problem that ranges across all departments and areas of campus life.

This apparent failure in communication and passing of valuable information and subsequently, the reduced accessibility to valuable information is the foremost problem obstructing the maintenance and growth of LMU's ES program and efforts. For example, when I first began my research for this study, I went onto LMU's sustainability website, `admin.lmu.edu/greenlmu`. I found that the website contained little information about LMU's ES programs and efforts, and what was available was outdated. I found a page that listed all faculty and staff involved with LMU's E2SE and GreenLMU. Many of the names listed on the page were individuals who were no longer at LMU. After, I found out about the outdated information the first time I visited these sites, I reached out to Rebecca Chandler, the Vice President of Human Resources at LMU, for information regarding staff and faculty who are involved with LMU's ES. She recommended I look at those same websites, and that although she knew they were outdated, that was best source for identifying potential ES individualseven for her.

Over my four years at LMU, and through my engagement with GreenLMU, CURes, and Facilities Management, I learned that LMU fails to promote university knowledge. Rather, only specific individuals have access or know key pieces of information. This inhibits the university's ability to grow collectively. Similarly, LMU does not make up for their lack of accessible knowledge through engagement and communication. This is why my following study is important. Through my study, I hope to better understand and explain the communication behavior around LMU's ES initiatives so that, should my findings be meaningful, LMU may begin to make changes to reunite the university in a common goal to progress and improve environmental sustainability.

Chapter 3

Literature Review

3.1 Organizational Structure and Behavior of the Firm

In the late 1900's, researchers began to study and model organizational structure. These studies looked at topics ranging from management, organizational behavior, culture, group dynamics, hierarchy and so forth. Perhaps the most accredited study and publication in this field is Richard Cyert's *A Behavioral Theory of the Firm* (1964) in which he discusses the affects and effects of individual and organization-wide behavior characteristics and their respective impacts on the organization as a whole. Of particular interest to my study are the generally concerted theories of Organizational Behavior (OB) and more recent theories about organizational attention[1]; and, more specifically, the recognized behavioral dynamics that supports these theories. A consideration for the aforementioned will provide the setting in which I can the platform my research questions, goals, and objectives.

[1]Researchers in this field are still very much questions, testing, and remodeling OB theories. However, there does exist enough consensus that many of the models established in the later 1900s are still being taught in universities today as a part of general Business curriculum ([29]).

Cyert (1964) grounds his research in the formal structure of firms. That is the highly visual publicized hierarchy based of position title and sanctioned decision-making authority on which firms segment and allocate tasks. Tree graphs visually depict formal organizational structures where those with the authority to approve decisions are few and at the top. Similarly, they clearly show segmentation of business tasks among employees and business units. As such, Cyert (1964) and later researchers suggest that business firms and organizations are coalitions in which there are distinct groups and groups of individuals that make up the whole [32].

The dynamics of a firm are both dependent upon and determined by the independent cultures of the coalitions and the interactions between the coalitions. Extreme incongruences between the coalitions can stifle the overall success of the organization. Most often, these detrimental disconnects are associated with goals and goal formations [32]. Since research exposed this problem nearly fifty years ago, many organizations have amended their strategic planning methodologies. However, with large organizations, even the best of solutions might fall short because of communication obstacles.

Rules and role expectations that form within and among the coalitions both self-imposed and prescribed by others largely affect the communication proficiency and organization attention. Miner (2006) affirms that the predominating structure and resulting expectations emerge from the interdependencies of the individuals and group actions and the distribution of competencies, values, and resources. If organizational culture places considerable emphasis on strict adherence to the formal hierarchy, involvement and attention to organization problems and projections may be forced. Alternatively, if the firm culture allows for more permissive and unsegmented structures, attention to and involvement in organizational decisions may be elective where status and title presents an individual with the right but not necessarily the requirement to participate in decision-making roles [32]. Consequently, it is obvious that there exist many layers and factors that influence decision-making and communication in organizational settings. Furthermore, solutions usually seek to adjust means of communicating among and between coalitions, while the general procedure for goal formation and decision approval remains the same: plan evaluation, examination of costs, demand, and objectives, and selection of appropriate model [32]. The more permissive and unsegmented an organizational structure, however, the more likely the aforementioned procedure is bound to complications.

The degree to which organizations emphasize adherence to the formal structure largely influences employee expectation and who pays attention to which company decisions. Particularly, in more unsegmented, permissive structures, status is considered a right to make a decision rather than to a requirement to participate in its evaluation and articulation. Consequently, many such organizations with permissive structures must learn to balance between individuals competing for a voice and too many apathetic, non-participatory individuals.

3.2 Comparing the Firm and University's Formal Organizational Structure and Behavior

The university is one such organization type that must continuously check and balance the attention and involvement of its employees and greater stakeholder community. Like traditional firms, universities host a hierarchical structure under which decisions travel and operate as a result of the information conveyed through a chain of interactions [26]. For example, in the following report, I limit the context of my study to activity pertinent to the progression of environmental sustainability initiatives at Loyola Marymount University. Environmental sustainability programs carried out at the university are theoretically decided upon through a filter of various departments and titles in hopes that they achieve or at least follow a complimentary trajectory for some greater university goal or objective.

After working in the Center for Urban Resilience and with Green LMU in the Sustainability Office, I learned that individuals anticipate that decisions will be made hierarchically through vertical processes of approval. Theoretically, at each level, the assigned reviewer contributes his/her own considerations and approval, assuring that the items first align with overall, long-term university goals and second, that the items are realistic and that the departments involved and the university has the resources, or can get the resources, for the program to succeed. Initiatives are supported by either internal or external demand, i.e. social trends, political policies, and supported studies, and the university must come up with and evaluate a plan to implement, manage, and maintain the plan, costs are examined, alternatives are evaluated, and the plan undergoes various reviews and reexaminations by both internal provosts and external organization, such as AASHE Star or U.S. Green Building Council. Anticipation and expectation, however, do not necessarily translate into actual performance, and communication does not strictly flow hierarchically. In reality, individuals communicate much more on an informal basis, so that how things are suggested to work is in conflict with how they actually do. As such, there exists an ongoing interest in understanding actual vs. expected organizational behavior particularly at more unautocratic organizations such as the university.

Cohen, March, and Olsen (1972) argue that even at the formal level, university decisions are not necessarily made with the same linearity as in the traditional firm. Concepts of firm culture, learning, experience, and sense-making all apply and fit with the university model as they do with firms. What sets the university apart from the firm, however, is that the qualities that make up the former are more pronounced and observable. The dynamic, exploratory, viva voce of the student body, staff and faculty encourage university heads to continuously address and question in what ways their efforts are building the most ideal and desired environment and community [32]. Consequently, universities have a permissive organizational structure with departments and coalitions that are so distinctive and generally disjointed that the university as a whole is characterized by problematic preferences, unclear technology, and fluid participation [11]. At the university, decisions are largely determined by who chooses to attend to a problem, their skill, capabilities, understanding and preferences, and available choicesnot necessarily an appointed team chosen to solve an existing problem. Particularly, Cohen et al. (1972) differentiate the university decision-making and operations from those of formal organizations by these three properties:

(1) **Departments and individuals across the university at all levels have inconsistent and problematic preferences and goals**: While many large firms experience similar difficulties, the liberal, largely horizontal structure of the university assuages any sort of strict standard. Even with clearly articulated and practiced community culture and expectations, the university is really a loose collection of ideas. In fact, much of a university's reputation is based on its ability to transgress boundaries and be at the forefront of new movements and trends, which requires a think-tank like culture.

The foremost distinction between traditional firms and universities is it fluidity. As a whole, faculty, staff, administration, and students maintain this fluidity through constant learning, questioning, and challenging expectations. This behavior is the quintessence of the university. Members of the university are encouraged to take risks and inspire innovation, operating on a basis of trial and error, accident, and experience. Rather than acting upon predetermined preferences, they discover preferences through action [11]. Furthermore, Cohen et al. (1972) suggest that because of its fluid structure, decision makers at universities are sensitive to increases in load, flight and oversight of problems depending on the time at which the problems are communicated, and because of the disconnect between various departments and coalitions, problems are not necessarily tracked and addressed efficiently appointed to or received by an available decision maker at the right time [32]. Accordingly, while universities might be grounded in a set of defined morals, the application of these values and the set of pertinent issues and the resources and decisions to achieve university leadership turn over almost as rapidly as its student body.

(2) **Not all individuals involved and who carry a stake in the university understand the processes, technology, and methods of operations**: Processes, technology, and methods of operation may be very technical and are largely constructed and determined by upper level management and administration. As a result, many individuals may never receive proper communication and instruction about the whole of a process, technology, or method beyond that which is pertinent to their specific task. Other individuals may dismiss the proper information themselves, believing it frivolous or secondary to work for which they are responsible. Consequently, the university as a whole experiences reduced coordination because of resulting discrepancies between the number of individuals who know and are savvy with the processes, technology, and methods of operations and those who cannot reiterate the technicalities and reasons for these constructions.

(3) **Lastly, individuals at universities are subjected to unpredictable movement of attention and devotion of time to various domains**: Even if members devote consistent time to an area, the ambitions and incitements of the university cycles with the movement of the students. This carries over into decision-making and subsequently, the programs carried out by the university. Thus, the problems, the work and the decisions that members address are a matter of choice behavior, constrained by time [11]. Only individuals who, time permitted, deliberately attend to particular problems assist in the generation of choices and ideas to resolve the problems or to create preventative measures.

The choice to attend to and participate with full awareness to a decision, however, it not always dictated by structured formal individual roles. Even if an individual's formal role requires their attention on a certain task or decision, there usually does not exist a strict method for accountability to prevent the individual from redirecting the majority of their time to their area of interest. Instead, the permissive university structure allows and encourages individual initiative to engage in activity of their interests and expertise under the assumption that each individual will manifest work that aligns with overall university values and goals. The main problem arises, however, when unpredictable devotion of time is accompanied by a reduced sense of accountability, permitting individuals to neglect reallocating, reassigning, or sharing important task information for the decision which they have abandon. This causes the university to experience a cyclical decline that begins with more structural slack, reduced observance of hierarchical order, less heterogeneity across the departments and university whole, and cause greater communication and goal alignment failure [32].

3.3 Understanding University Organizational Structure Through Informal Networks

Through is particularization of university structures from traditional firm structures, Cohen et al. (1972) articulated the need for a different approach to the understanding of the interactions and decision-making process occurring at unsegmented, permissive organizations. Since universities are unlike traditional firms, theories of organizational behavior and choice behavior associated with traditional firms and systems are insufficient. A relatively new field called Social/Organizational Network Analysis (SNA/ONA) suggests that network analysis and the study of organizations' relationship networks provide insight into individuals' performance, importance, and weakness within a company, and consequently, the impact of individuals' interlocking behavior on organizational success. Essentially, SNA captures information about organizations' invisible, informal network built from dyadic social relationships and through which information travels unique to those relationships rather than route prescribed by the organization [41]

While researchers have long recognized the importance of informal networks within organizations, they are just recently beginning to understand the depth of its impact on organizational behavior and decisions [45]. In 1964, Peter Blau studied the associations of individuals and their tendency to build social structures and interpersonal relationships. Later, researchers applied Blau's work and insight to the context of the work place. As a result, the OB field began to discuss more the impact and effects of social environments of the work place. In 2007, McKinsey & Company exclaimed that companies who attempt to apply OB theories and changes to their organization system will fail 2/3 of the time if they chose to ignore the invisible networks and the informal social relationship. Informal social relationships support social capital, knowledge transfer, organizational learning, communication, leadership, and power [45]. The informal structure, opportunities, and behavior amalgamate to create an organization's culture and organizational development and aid to an organization's overall competence and can illicit superior collaborative learning [37]. In order for success to actualize, however, social capital and knowledge need to be transformed into organizational knowledge and thus reduce dissonance across the university [8]. Consequently, these relationships help better identify the truly efficient distribution of resources, organizational attention to different domains, and construction of social capital that lead to actualization of success than those portrayed by the formal structure [5].

In fact, adherence to and strict following of formal organizational structure, obligations, behavior, and decision-making processes at both traditional and permissive organizations is active only through expectation. Instead, much of organizational success is formed unintentionally through the non-specified interactions of the employees. In other words, members of organizations expect decisions to be made following an explicit structure whether that structure is linear or non-linear. In reality, however, communication and development of decisions and initiatives tend to follow and evolve through organization's informal social networks especially if the organization structure is more permissive and free flowing, such as a university [37].

Bekessy, Samson, and Clarkson (2007) acknowledge informal behavior is particularly influential in the context of commitments to environmental sustainability by universities. Successful adoption of sustainability into university culture and practices requires support from all areas and department of the university, and especially leadership high-level university officials [40]. Thompson and Green (2005) promote ground-up strategies suggesting that sustainability need not be a priority. Instead, they hold that university students and small-scale club activities may play a role in the redirecting faculty, staff, and administrative attention through cooperative demonstrations and awareness rising. When the student body is itself unified, their persistent demands, propagated ideas, and proposed choices for change force decision makers to respond. Though, even with a unified student voice, the permanent, lasting success of environmental sustainability programs requires the engagements of faculty, staff, and administration whose authority can concede or reject campus cultural and behavioral changes [40]. Furthermore, the general four-year cycling of students leaves programs vulnerable to phasing out with the turnover of the students and lets university decision makers bypass environmental sustainability efforts and direct their attention elsewhere.

Instead, lasting environmental sustainability success requires the support of decision makers, faculty, staff, and admin because they have the authority to change "institutional arrangements" that enable systemic change [18]. Still, exclusively top-down strategies fail to accomplish a long-lasting materialization of environmental sustainability on campus unless there exists the right culture conditions and the support of a majority of the governed body, and as McKinsey pointed out, without recognition of the university informal structure and the ensuing behaviors, both top-down and ground-up organizational strategies will fail. Instead, by adjusting university decision processes to fit the informal network, universities can better understand how to build collaboration and reciprocal support between small-scale university activities and upper-level decision-makers. This dynamic teamwork will also create a system that better upholds individual and departmental coordination and accountability and allows for greater cross-functional social capital permeation [40].

While it has recently become evident that understanding an organization's informal network is critical to long-term success, doing so requires significantly more effort because, unlike organizations' formal hierarchical structure, the informal structure is not explicitly dictated. This is the particular role of SNA. Generally defined, Social Network Analysis is a "set of methods . . . which are specifically geared towards an investigation of the relational aspect of the structures" [41]. Instead, these methods use relational data rather than attributed data to say something about the organization or network being studied. Informal networks are built through the interactions and associations of individuals separate of their hierarchical status and formal role within an organization [45]. People are drawn to others through interpersonal attraction, like schedules, shared interests etc. Furthermore, informal interactions provide individuals with a greater sense of belonging, control, and worth so that many individuals seek out such informal relationships for their own personal benefit. In other words, many form informal relationship with those within the network who not only hold common interests, but also who through trust, accomplishment of results, and efficacy maximize their own individual utility, benefits, success within an organization and potential for reciprocity [45].

SNA uses surveys particularly designed to capture these inter-personal relationship data based on commonalities and utility maximization behavior. They also make visible previously imperceptible characteristics about the inter-organization relationships and the de facto design of individuals' and departments' interaction such as which individuals are important in connecting different departments or whether certain departments are siloed, why they might be siloed, and the potential effect on the whole of the organization [14].

There are different models and methods for analyzing SNA data depending on the core research question and purpose of the study. When applied to business and organizational context, SNA tracks the diffusion of information from person to person in order to build the organizations informal network. To look simply at who speak with whom, however, provides insight only into the general social network of the organization. Many times, this can be useful. However, it is usually advantageous for researchers to identify a framework in which to observe particular communication.

For example, say you wanted to study trust within a company. You test to see with how many different individuals each person spoke. If certain individuals withhold from speaking members of their team, perhaps they do not have a lot of trust in their coworkersThe use of these methods depends on the availability of the specific relationship data of interest [41], which is in part dependent upon the design of the surveys used. In the example used above, if the survey questions and the SNA methods are not particular chosen to test trust behavior within the organizational network, the data is subjected to false analysis. One might look at the data and suggest that the isolated individual is the trustworthiest because they do not speak with other people.

Furthermore, SNA researchers benefit from applying general network analysis to their constructed informal networks as well. Novak et al. (2011), for example, provide five different circumstances that require different analysis of SNA data. General network measurements also illuminate previously invisible and informative network characteristics. For example, Novak et al. (2007) give these five examples where they applied SNA: 1. Cross-Functional collaboration, 2. Trust and communication, 3. Teamwork detected only at the top, 4. Change in culture, and 5. Revealing Patterns of Effectiveness. To study cross-functional collaboration, researchers use data collection about information diffusion between individuals to form the informal network, but they might analyze the network graph by identifying how many connections exist between individuals and departments. (I will define and explain fundamental network properties and characteristics used in network analysis in the following section).

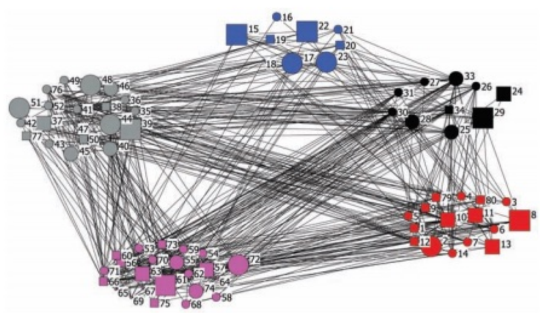


Figure 3.1: Example of network density between departments from Novak et al. (2007)

As noted before, SNA uses the "diffusion" of information to build and analyze organizational networks whether a traditional firm or a university. The concept of studying groups of people through network analysis and the flow of information is not nearly as old as the game of telephone, but it is also not so new. The rise of social media in the early 2000's however, greatly encouraged an expansion and refocusing of the field. This is when researchers began to network analysis and the flow of information with organizations. Banerjee, Chandrasekhar, Duflo, and Jackson (2013) added to the field by advancing a new method called "Diffusion Centrality." These individuals did not create this idea of diffusion centrality, but instead were one of the first to successfully conceptualize, apply, and explain the new metric. "Diffusion" simply means the "transmission," "flow," or "dissemination" of the form [information] of network. "Centrality," on the other hand, is more complicated concept [17]. I will discuss this in much greater detail in the chapter, Breakdown of Mathematical Models and their Supporting Theories, but for now it is sufficient to understand "centrality" as "a vital, critical, or important position" principal to the "diffusion" of information across a network or group of people. Essentially, in SNA, Diffusion Centrality identifies which persons are information hubs and have the greatest influence on that information reaching the rest of individuals within the network.

3.4 Communication Centrality and Diffusion Centrality as Arithmatic Methods for Studying Informal Organizational Networks

On the surface, it may seem relatively easy to identify a person or persons in an organization or group who may be the most adept at transmitting information to others within that organization or group. In actuality, it is more complex, and has been the subject of detailed scientific studies and mathematical representation to achieve the identification of individuals with some degree of certainty. One of the first studies to make use of and advance the idea of diffusion centrality is Banerjee, Chandrasekhar, Duflo, and Jackson (2013)'s *Diffusion of Microfinances* and their subsequent study, *Gossip: Identifying Central Individuals in a Social Network* (2016). Through their work in 2013 and 2016, Banerjee et al. proposition that network communication travel and resulting structural effects may be predicted by understanding individuals' centrality within a network through their communication behavior, preferences and utility maximization choices.

In 2013, Banerjee et al. detailed their study on communication behavior in *The Diffusion of Microfinance.* They learned that the adoption of new microfinance information and programs in India was substantially more predictable when information was seeded through communication central individuals than through individuals identified by other models which use other qualifications to test how influential and central individuals are within an organizational system or network. In other words, they use individuals' communication behavior to understand network effects as a whole overtime. This idea puts into mathematical practice what scientists like Huning and Brown have articulated about informal organizational relationships, and applies the idea of centrality as a newer metric to Social Network Analysis.

To calculate and quantify communication behavior, Banerjee et al. (2013) developed a Communication Centrality model, using estimated parameters specific to their study. They seeded information through individuals in their network to then identify those who were communicatively central and most influential in dispersing information about the network. They then used this data to forecast who would and who would not eventual participate in their study's new microfinance program if the details of the program were seeded through the communicatively central individuals. This model, essentially, allows Banergee et al. (2013) to discover the realities of communication and information transfer within a closed network, such a busy organization. Their Communication Centrality model, however, relies on computed parameter estimates found through simulated data specifically related to their study context in India. So, to generalize their Communication Centrality model, they developed a proxy called Diffusion Centrality.

Diffusion Centrality, instead of imposing characteristic parameters, assumes that every person in the network shares information with the same probability, and thus uses network properties to determine whether individuals will pass along information. As such, the model is mainly concerned with diffusion centrality as it results from network connections, where the probability that individuals are connected with others within the network is subjective to each individual. As such, the Diffusion Centrality model is much more generalizable than Banerjee et al. (2013)'s Communication Model. And with this model, researchers may more easily garner information about organization's informal coalitions as proposed by Huning et al. (2015).

Banerjee, Chandrasekhar, Duflo, and Jackson later revisited their data to test whether they could get the same diffusive centrality result strictly from relationships mentioned in their study's survey without first seeding and tracking information throughout the network. They called this model, Network Gossip. Through their work in *Gossip: Identifying Central Individuals in a Social Network*, Banerjee et al. (2016) could not reject their hypothesis that their newly developed Network Gossip model pinpoints diffusively central individuals for information communication. In other words, the researchers turned their analysis on its head by starting "backwards," requesting individuals to answer survey questions in which they nominate others within the network according to different communication characteristics. In doing so, the researchers actually advanced the concept of Diffusion Centrality.

The uniformity of the questions among all survey participants increases the likelihood that individuals communicate with the same probability, and thus reduce the deficiencies caused by an assumed equal communication probability. As a result, this model better identifies network characteristics by recognizing communication behavior. Accordingly, I posit that this model may be the second step in understanding the impact of informal organizational networks through its ability to identify the flow of social capital and information, and find, as Brown (2000) urges, who within the network helps to transform that information into resonant knowledge across the university. Should Banerjee et al. (2016)'s Network Gossip model be adapted appropriately, applying the model to a university context will uncover a knowledge of how individuals interact and through which relationships individuals perceive as beneficial, which will help develop better projections of informal organizational behavior and lead to greater overall organizational success. Specifically, this model will allow me to discern informal communication and collaboration behavior around LMU's environmental sustainability initiatives.

Chapter 4

Research Questions

The objective of my research is to better understand the informal communication network about environmental sustainability at LMU. I seek to gain insight by examining the environmental sustainability network structure within the university as it is perceived by those involved. Networks can be described in general, common terms, but scientists and researchers have studies and developed various network models that depict networks from a scientific standpoint both graphically and mathematically for the purpose of understanding how the networks function, can be manipulated, and made more efficient. In other words, a network in its simplest form seem relatively superficial, but it can be geometrically more complex and sophisticated, and can have a variety of applications.

One aspect of my thesis is to explore the graphical and mathematical depth and application of Banerjee et al. (2016)'s new network model called the Network Gossip model. I plan to test the reproducibility of this model, using a single university as the context and setting for my research rather than multiple villages. Additionally, I will correlate my understanding of this model with how LMU may best advance its environmental sustainability goals. Since the method used in the Network Gossip model requires considerably less time and resources to gather the pertinent data than the Diffusive Centrality model. This study explains various terms, such as diffusion centrality, eigenvector centrality, and network gossip centrality, as well as others, that are key to the design of the study and the mathematical formulas used therein. In the following section, I breakdown and explain theses terms and formulas and their significance to the design and understanding of my study. Successful adaption and application of the Network Gossip model to a university context will support and encourage greater understanding of informal, social organizational networks and provide a platform of understanding from which to reevaluate university communication and behavior. I hope that any empirical data and results will be beneficial for those interested at LMU and motivate the reintroduction of the E2SC meetings (or similar assembly), a collective effort towards the advancement of LMU's environmental sustainability initiative, and an overall attempt to reduce the campus' carbon footprint.

There are several key elements to understanding the analysis of a network. The simplicity or complexity of these elements matches the simplicity or complexity of the desired analysis. Of particular interest in more in-depth network analysis is the conceptualization of a network in a multi-dimensional space. That is, how the network performs and changes over time. Researchers apply these different analyses to, in the end, determine how and to what extent each node (or participant) in the network influences both the flow of information and the impact or affect that communication has on the ultimate output of the network. Among the elements are various manifestations of the centrality of a node (e.g. degree; closeness to other nodes; decay (i.e. weighting to determine dissipation); betweenness; diffusion; and other concepts specially names after the authors of the particular analysis.) The following segments of my thesis will discuss these elements and their different arrangements in detail, which can be expressed either graphically or through mathematical formula.

In my research, I examine the formal organizational network structure around environmental sustainability at LMU by departments and within departments to build an expected typical information communication path that I then compare to an informal network structure constructed by the Network Gossip model. During the analysis, a few key questions drove my interpretations and thinking:

- What, if any, are the differences between the formal and informal communication network structures about LMU's environmental sustainability?
- Do the relationships identified through the Network Gossip model accurately depict individuals relationship insofar as they relate to communication about LMUs environmental sustainability?
- What if anything do the constructed networks tell about LMU's environmental sustainability efforts and LMU in general?
- How might the identified whole network properties and specific nodal properties help individuals and departments advance the universitys environmental sustainability efforts?
- What are the university-level outcomes of the current communication network structure?

Chapter 5

Breakdown of Mathematical Models Used in this Study and Their Supporting Theories

> "Acquaintances, in sort, represent a source of social power, and the more acquaintances you have the more powerful you are." Malcolm Gladwell,
> *The Tipping Point:*
> *How Little Things Can Make a Big Difference*

In the following pages, I seek to explain in plain English the essence of the study, including the actual mathematical formulas and related theories presented in *Gossip: Identifying Central Individuals in a Social Network* [2].

There is a handful of definitions that are fundamental to understanding the scope of the following study and the basis on which I make my observations and draw conclusions regarding the interactions among LMU departments, faculty, staff, and students regarding the school's environmental sustainability efforts. The definitions of various network properties are set forth below. Again, the terms used vary depending on context. Therefore, I will include both the general definition and if necessary, any other terms that I will use interchangeably to reference the same concept. Also, throughout the paper I will use i and j as general terms that exist within the network, $G(N, A)$, which consists of a finite set, N of nodes and a finite set, A, which connect pairs of nodes. An edge connecting nodes $i \in N$ and $j \in N$ will be denoted as (i, j) where $i \neq j$. The structure of the network itself allow us to test and identify node characteristics and say something about node relationships and the network as a functioning system. I will continue by addressing more complicated network measurements, specifically those related to centrality. From there, I will begin to explain the mathematics behind Banerjee et al. (2016)'s Diffusion Centrality and Network Gossip models and rationalize the use of these SNA models in my own study.

Throughout this explanation in "Breakdown of Mathematical Models Used In This Study And Their Supporting Theories," all work is that of Banerjee et al. (2016) unless otherwise stated. I use various terms such as "we" and "us" to make the material more engaging, and do not claim any of the work as my own. In the following I simply attempt to breakdown and clarify the work done by these mathematicians, and demonstrate how this information and mathematical depiction are important to the analysis of my study.

5.1 Network Properties and Graphical and Mathematical Depictions

In the following glossary of terms, the definitions presented come from Jackson (2008) unless cited otherwise.

Network – Taken from the Oxford Dictionaries, the most general definition of a network is, "an arrangement of intersecting horizontal and vertical lines." This definition does a good job of explaining what networks look like visually. However, to truly comprehend the idea of networks depends on context. We can begin to add depth to the definition by recognizing networks as "a group or system of interconnected people or things" [16]. Add the idea of relationships to the previous definition. So networks not only define a visual, physical structure, but also less perceivable characteristics. From here, we can begin to base our definitions on context. (i) for mechanics, networks can be defined as "a system of computers, peripherals, terminals, and databases connected by communication lines," [15]; (ii) for biology, "any group of neurons that conduct impulses in a coordinated manner, as the assemblages of brain cells that record a visual stimulus" [17]; and (iii) even social networks, "a group of people who exchange information, contacts, and experience for professional or social purposes" [16]. All of these definitions are slightly different largely because different fields of study use different terminology to express similar concepts. What makes these definitions even more meaningful than even the second definition is the perception of communication and interaction among theses relationships. This thus implies that these are behavioral relationship, which provides for much more interesting, important, and complex research.

Network Graph[1] – most analyses depend on the construction of a network graph to represent and illustrate relationships. Network graphs are denoted as (N, g) where N is the number of nodes in the sample and g is the corresponding n x n matrix. g is formally called a transition matrix where a marked cell at (i, j) indicates a directed,one-way connection from node i to node j.

Node – a single element in a network where elements are typically identified by a letter such as i, j, and k, or a number; vertex; individual.

Link – a connection between two nodes; edge.

[1]All subsequent definitions in section come from Jackson, Matthew (2008), Social and Economic Networks. Princeton University Press, Princeton, NJ. p3-34.

Direct Link – a single direction link; for example, a link between a citizen and the president where the citizen can identify a link to their president, but the president may not be aware of his link to the citizen.

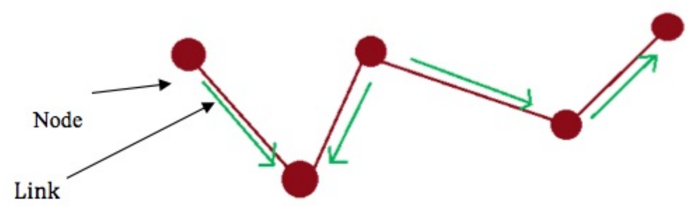


Figure 5.1: Arrows show the direction of the relationship between the linked nodes.

Indirect Link – a link with no specific direction, usually used to represent a reciprocal relationship; for example, a friendship where both individuals are aware of the other; arc.

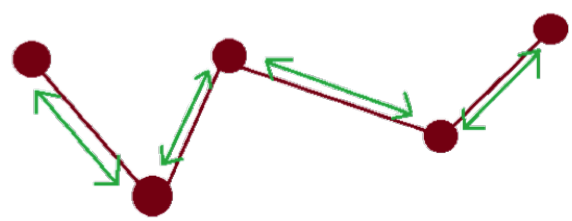
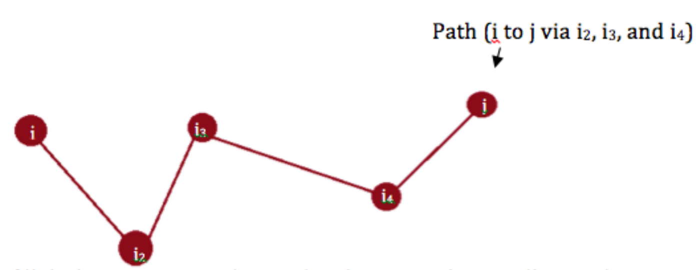
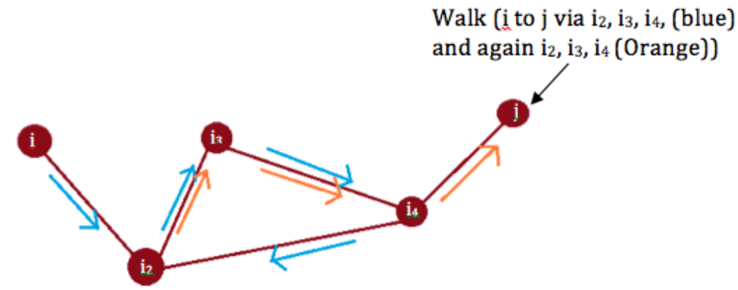


Figure 5.2: Arrows show the direction of the relationship between the linked nodes.

Path – the links between two nodes i and j, where each node is distinct.



Walk – a sequence of links between two nodes i and j where some intermediary nodes may be revisited more than once.



Geodesic – the shortest path between two nodes i and j.

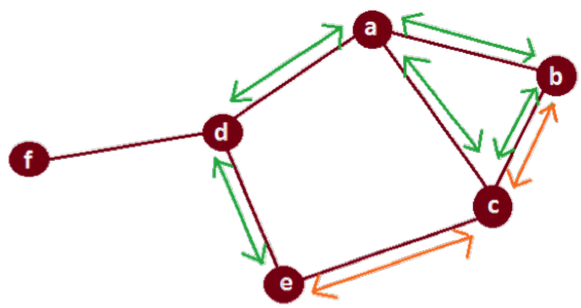
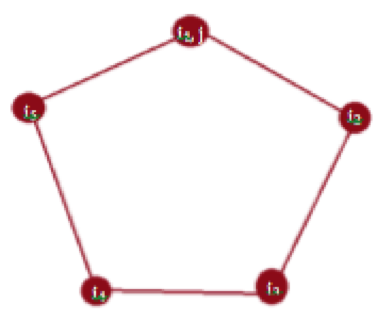


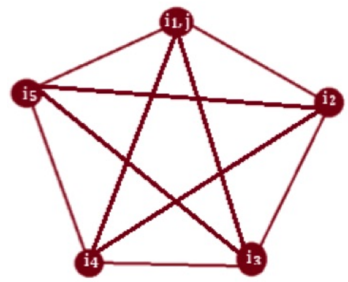
Figure 5.3: Some of the possible paths between node e and node b; Green arrows show various paths reguardless of length. Orange arrows shows the geodesic, [e,c,b], between e and b.

Cycle – a walk that starts and ends at the same node so that $i_1 = i_k$ and only $i_1 = j$ appears more than once.



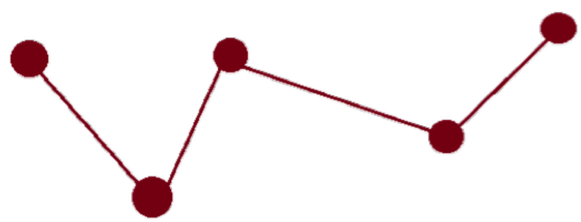
Connected Network – when the network does not contain subgraphs and every node is connected by some walk in the network

Complete Network – when every node in the network is linked to every other node in the network.



Neighborhood – all the nodes surrounding a particular node of interest by a distance of 1; neighborhood of node $i = N_i(g) = j$: $g_{ij} = 1$.

Distance – length (ℓ); number of links between two nodes.



Degree – the number of links that involve a node; a count number of a nodes immediate (1st degree) neighbors.

Density – the relative fraction of links that are present in the whole network[2].

Diameter – largest geodesic distance existing in the network such that any node may reach any other node by a path of no more than the largest geodesic distance.

[2]Potential Connections in Network: ($PC = \frac{n(n-1)}{2}$); Network Density: $\frac{ActualConnections}{PotentialConnections}$ ([39]).

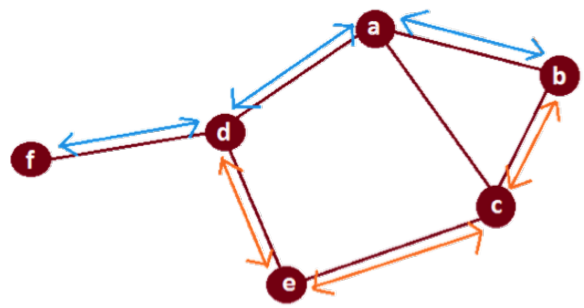


Figure 5.4: The longest geodesics are of length 3 [f,d,a,h] shown in blue & [d,e,c,b] shown in orange; All other geodesics in this network graph are shorter than length 3.

Characteristic Path Length – the average over geodesics in a network bounded by the diameter.

Connected Network – when the network does not contain subgraphs and every node is connected by some walk in the network. (See network shown for "Complete Network").

Subgraph – a nonempty (at least one node) subnetwork within network (N, g); component.

Adjacency matrix – A matrix representative of links between nodes in which two nodes can either be linked, 1, or not linked, 0, $A \in \{0,1\}^{nxn}$. If a link between two nodes a and b is direct, then elements $X_{ab} \neq X_{ba}$. Conversely, if a link between two nodes a and b is indirect, then elements $X_{ab} = X_{ba}$. In the example below, we are assuming indirect relationships between the nodes. Likewise, for the remainder of the paper I will use this assumption unless otherwise stated.

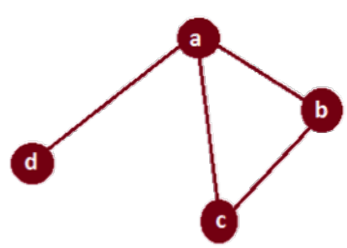


Figure 5.5: Simple Network Example

	a	b	c	d
a	0	1	1	1
b	1	0	1	0
c	1	1	0	0
d	1	0	0	0

Figure 5.6: Adjacency matrix corresponding to previous figure.

5.2 Network Properties – Centrality Models

The terms above define different elements that found a network. Separately, these definitions allow us a basic understanding of network graphs and how we may visualize network connections. By integrating these network parts, we can begin to build a more comprehensive understanding of the network graph.

More complex network properties and network models use variations and combinations of the elements defined above to tell something about the nodes individually and as an aggregate. One such concept is centrality. Centrality is a measure of how important a node is within a network. Centrality measures can have "local" and "global" characteristics. Local centralities explain a node's structural position, qualities and outcomes, within a network system. Global centralities, on the other hand, define node characteristics by their influence on system-level outcomes [38].

[3] For this study, I focus strictly on Bonacich Centrality with a positive attenuation factor. That is, the weights given to the linked nodes ranges between 0 1, and diminishes as the nodes distance away from the node of interest increases. Conceptually, this method states that a nodes power within a network is determined by how well connected it is to other well-connected neighbors. Conversely, however, researchers might consider a node to be more powerful the more that other nodes dependent upon it to interact themselves within a network (similar in a sense to how Betweenness Centrality looks at how critical a nodes is in connecting others). Thus, this method uses a negative attenuation factor where the weight given ranges between -1 0. Consequently, a node that is less well-connected within a network will have a greater impact on the node of interests power within the network ([7]).

Model	**General Model Equation**	**Strengths**	**Weaknesses**
Degree Centrality	$d_i(g) = \sum_i g_{ij}$	Look at the number of neighbors linked to each node. Quickest measurement to calculate to get a brief snap shot of the network structure	More simple and therefore the least information. Subject to many omissions [43].
Closeness Centrality	$\frac{(n-1)}{\sum\limits_{j \neq i} \ell(i,j)}$	A slightly more complete measurement than Degree Centrality as it positions a node within the whole network structure.	Like all centrality measurements that are summed over the whole network, if there exists disconnected nodes within the network, the measurement value equals infinity, and therefore, must omit the disconnected nodes from the network to obtain a value. This has the potential of obscuring the interpretation of the network structure [36].
Decay Centrality	$\sum\limits_{h \neq a} \delta^{d(a,h)}$	Provides a deeper evaluation of Closeness Centrality, by weighting the nodes that extend out from the node of interest according to their distance away from that particular node.	Dependent upon an exact topology of the network structure. Can be very difficult to calculate for complicated networks and network graphs [46].

Table 5.1: Centrality models

Model	**General Model Equation**	**Strengths**	**Weaknesses**
Betweenness Centrality	$\sum_{j \neq k : i \in \{j,k\}} \frac{P_i(jk)/P(jk)}{(n-1)(n-2)/2}$	A unique measurement that looks at a node importance as an intermediary node rather than its influence as an end node [36].	Typically, Betweenness Centrality only looks at the fraction of shorts paths on which the node lies in order to determine the nodes central importance. This however, neglects all other potential paths as options of the spread of information, disease, etc. [33].
Katz-Bonacich: Katz Centrality	$P_i^K(g) = \sum_{j \neq i} g_{ij} \frac{P_i^k(g)}{d_j(g)}$	Reveals how nodes with the same degree and vary in network importance and that a node's influence is dependent upon the degree of its neighbors and neighbors' neighbors etc. [22].	Requires complex calculations and requires one to approximate an appropriate ∞ for the network system. Therefore, this model is usually only applied to simulated data.

Table 5.2: Centrality models continued...

Model	**General Model Equation**	**Strengths**	**Weaknesses**
Katz-Bonacich: Bonacich Centrality[3]	$P^{K^2}(g,q) = qg \cdot 1 + q^2 g^2 \cdot 1 + ... + q^k g^k \cdot 1$	Like Katz Centrality, Bonacich Centrality reveals how a node's influence is dependent upon that of their neighbors and neighbors' neighbors etc., but also weights each linked node whereby the weight is taken to the power of the nodes distance from the node of interest, and thus accounting for the node's lesser role on the node of interest's power within the network [22].	Requires complex calculations and requires one to approximate an appropriate ∞ for the network system. Therefore, this model is usually only applied to simulated data.
Eigenvector Centrality	There does not exist a general equation form for Eigenvector Centrality. Instead, $v^{(R,1)}$ represents the largest right-hand vector, and λ_1 is its corresponding largest real eigenvalue.	–	–

Table 5.3: Centrality models continued...

5.3 Network Properties – Centrality Models – Degree, Closeness, and Betweenness Centralities

Degree, Closeness, and Betweenness centralities are position-based centrality measures of a node that is they are calculated based on the node's connectedness and relative location within a network. To help illustrate the three centrality measures listed above, consider the following situation:

Imagine you are back in college, going home for Thanksgiving Break. You find out that your family is having only a small Thanksgiving Day celebration with 10 of your favorite family members. This includes, you (u), your older brother, Bill (b), your younger sister, Alice (a), mom (m), dad (d), your aunt, Ellen (e), on your dad's side, your uncle, Fred (f), their two boys, George (g) and Harrison (h), and your grandfather, Carl (c), on your mom's side. Limiting our view to this story, these individuals make up our social network. Links within this network represent immediate relations. This is so our network has some depth. If we were to link individuals base on whether or not they know each other, we would end up with a complete network and the number of the centrality values for each node would be the same. So, this is our network and correlated adjacency Matrix B:

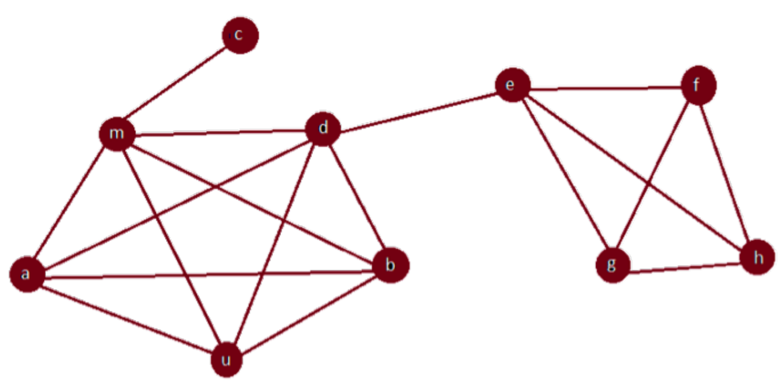


Figure 5.7: Family network graph.

	a	b	c	u	m	d	e	f	g	h
a	0	1	0	1	1	1	0	0	0	0
b	1	0	0	1	1	1	0	0	0	0
c	0	0	0	0	1	0	0	0	0	0
u	1	1	0	0	1	1	0	0	0	0
m	1	1	1	1	0	1	0	0	0	0
d	1	1	0	1	1	0	1	0	0	0
e	0	0	0	0	0	1	0	1	1	1
f	0	0	0	0	0	0	1	0	1	1
g	0	0	0	0	0	0	1	1	0	1
h	0	0	0	0	0	0	1	1	1	0

Figure 5.8: Matrix B corresponding to Family network graph above.

Imagine also, that communication can only flow from person to person if they are linked. Yes, Thanksgiving Dinner here would be like one bizarre game of telephone, but this assumption will help clarify other points. Also, like at any stereotypical family reunion, people gossip, share big news, and incite drama. This time, your brother Bill announces to the family for the first time that he got engaged and will be married in March.

Degree Centrality, a local centrality characteristic, is the simplest of the measurements. It is equivalent to the number of neighbors a node has. We can identify each individual's centrality by summing the number of nodes to which each is linked within our network. Let the general notation of this Degree Centrality summation be represented by $d_a(B) = \sum\limits_a B_{ah}$ where $\sum\limits_a$ tells that we are finding the degree of node a by summing the number of links extending from a in matrix B. Looking at our family network in Figure 5.6 and using Matrix B in Figure 5.7, we can find each individual's degree distribution by counting the number of "1"s in each row.

$$d_{Alice}(B) = \sum_a B_{ah} = 1+1+1+1 = 4 \qquad d_{Dad}(B) = \sum_d B_{ah} = 1+1+1+1+1 = 5$$

$$d_{Bill}(B) = \sum_b B_{ah} = 1+1+1+1 = 4 \qquad d_{Ellen}(B) = \sum_e B_{ah} = 1+1+1 = 3$$

$$d_{Carl}(B) = \sum_c B_{ah} = 1 = 1 \qquad d_{Fred}(B) = \sum_f B_{ah} = 1+1+1 = 3$$

$$d_{You}(B) = \sum_u B_{ah} = 1+1+1+1 = 4 \qquad d_{George}(B) = \sum_g B_{ah} = 1+1+1 = 3$$

$$d_{Mom}(B) = \sum_m B_{ah} = 1+1+1+1+1 = 5 \qquad d_{Harrison}(B) = \sum_h B_{ah} = 1+1+1 = 3$$

The calculations above, however, only calculate degree distribution and simply identify a node's neighborhood. In order to assess degree as a centrality measure, we must compare each node's degree value relative to those of the other nodes in the network (Robin, 2015). Instead of looking at a node's degree in isolation, we are interested in each node's degree relative to that of the other nodes in the network. It is this comparison that exposes, in one-sense, an individual's prominence and influence in the structure of the network. As a result, Alice (4), Bill (4), and you (4) are more degree central than Ellen (3), Fred (4), George (3), and Harrison (3). Carl (1) is the least degree central and Mom (5) and Dad (5) are the most degree central of all. This means that if your brother thought that the best way to share the news of his engagement was to tell the family members connected to the greatest number of individuals, he would be most successful speaking to your mom and dad instead of you and your sister.

Two other measurements include, Closeness Centrality and Betweenness Centrality, which are also local characteristic measurements. Closeness Centrality tells how close, distance-wise, one node is to any other node in the network [25]. Recall that distance in network analysis refers to the number of links between two nodes, and the shortest distance or path between two nodes, is called a geodesic. Both of these properties are used to find a node's Closeness Centrality. To find node b, Bill's, Closeness Centrality we look at the inverse of his average distance. The average distance is found by summing all geodesics extending from node b and dividing it by $(n-1)$, the number of nodes in the network minus 1 since node b is not looped to itself.

$$\frac{(n-1)}{\sum\limits_{h \neg b} \ell(b,h)} = \frac{(10-1)}{\sum\limits_{h \neg b} \ell(b,a)+\ell(b,c)+\ell(b,u)+\ell(b,m)+\ell(b,d)+\ell(b,e)+\ell(b,f)+\ell(b,g)+\ell(b,h)}$$

$$= \frac{9}{\sum\limits_{h \neg b} 1+2+1+1+1+2+3+3+3} = \frac{9}{17} = .5294$$

Below are additional Closeness Centralities for other individuals in the family network:

Alice: $\frac{(n-1)}{\sum\limits_{h \neg a} \ell(a,h)} = \frac{9}{17} = .5294$ Dad: $\frac{(n-1)}{\sum\limits_{h \neg d} \ell(d,h)} = \frac{9}{13} = .6923$

Carl: $\frac{(n-1)}{\sum\limits_{h \neg c} \ell(c,h)} = \frac{9}{24} = .375$ Harrison: $\frac{(n-1)}{\sum\limits_{a \neg h} \ell(h,a)} = \frac{9}{21} = .4286$

In the general equation, the node named in the right of the pair and on the left of the inequality simply serves as a place holder, representing every other node in the network. Notice how when we expand the summation "h" is replaced by the appropriate end node. Likewise, $h \neq b$ just states that we look at every node in the network so long as it is not equivalent to b. So, we do not consider node b against itself.

This sort of calculation allows us to look at a nodes' distance from others as a ratio to the total number of nodes within the network. Nodes with a relatively greater number of geodesics extending from itself to all other nodes in the network receive a greater closeness centrality value than those whose path lengths between itself and all other nodes are longer than the network geodesic. In our family network example, Figure 5.6, we see that of the values calculated, Dad has the greatest Closeness Centrality, and, in fact, he is the most closely central in this network. If a node were directly linked to every other node in the network, then the sum of the node's geodesics would simply be equal to its degree and its Closeness Centrality would equal 1.

The Closeness Centrality measurement described here treats each link in the geodesic paths equally, suggesting that for Dad, node d, a link between himself and Mom, node m, is as meaningful and useful to him as the link between Mom and Carl, node c. Realistically, however, this is not always the case. Your dad maybe able to pass along information via your mom to convince your grandfather to provide funds for your brother's wedding, but your grandfather might be much more willing to do so if the request came from your mom directly. In such a case, we want to assign the links weights to exemplify just how useful they are to a particular node. This is called Decay Centrality.

Decay Centrality assigns weights, δ, to the link by considering a parameter $0 < \delta < 1$. The farther away a node is to node a, the greater the length between two nodes and smaller the δ; the closer a node is to node a, the shorted the length between the two nodes and larger the δ (Jackson, 2008). As δ decreases and approaches zero, the more proportional Decay Centrality becomes to Degree Centrality. As increases to 1, the more Decay Centrality depicts the size of the component in the network in which node a lies [46]. These weights are multiplied respectively to every link in the geodesic. The calculations to find Decay Centrality from this point are the same as Closeness Centrality[4].

Betweenness Centrality measures at how important a node is in connecting two other nodes in the network. Generally, a node i's Betweenness Centrality is the number of geodesics it lies on in the network (Jackson, 2008). The closer to a node's betweenness centrality value is to 1, the more essential it is in bridging a path between many pairs of nodes in the network. In other words, many of the shortest paths between nodes in the network include i in the path.

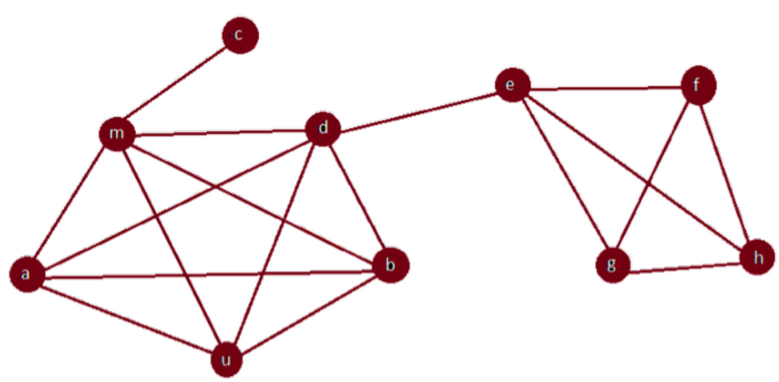


Figure 5.9: Family network graph.

[4]Decay Centrality: where $0 < \delta < 1$ is the decay parameter and $d(\cdot, \cdot)$ is the geodesic between the two nodes. The value chosen for the decay parameter depends on context. For example, if measuring radioactivity, the decay parameter is usually .5, the half-life of the material at a given time period [47].

Above is our family network from Figure 5.8. If we wanted to calculate Moms Betweenness Centrality, our equation would look like this:

$$\sum_{c \neg g : m \notin c,g} \frac{\frac{P_m(cg)}{P(cg)}}{\frac{(n-1)(n-1)}{2}}$$

To understand what this equation shows, let's begin by understand which terms are changed as we expand out the summation. c and g are place holders and can be replaced by any combination of end nodes, excluding the node of interest, node m. In fact, $(n-1)(n-2)/2$ in the denominator represents the number of possible end node combinations that we can make from our network sample size. $(n-1)$ and $(n-2)$ represent the end nodes of the path where $(n-1)$ is the first end node chosen from the sample and $(n-2)$ is the second end node chosen from the sample. Under the sigma,$c \neq g$ is stating that our two end nodes cannot be the same as they would be if our network were a cycle. $m \notin \{c, g\}$ just says that our node of interest, node m, cannot be an end node when calculating its Betweenness Centrality.

$P_m(cg)/P(cg)$ in the numerator is the ratio of geodesics between two other nodes in the network on which node m lies to the total number of geodesics for that linked pair. P_m simply expresses that we are looking at a geodesic path on which Mom, node m, lies. $P_m(cg)$ means that this path is between Carl, node c, and George, node g. $/P(cg)$ tells us to divide by the total number of geodesics between any nodes c and g.

For instance, Mom lies on the shortest path between you and Carl, $[u, m, c]$, and because there only exists one geodesic path between you and Carl, this ratio is equal to 1.

$$P_m(uc)/P(uc) = 1/1 = 1$$

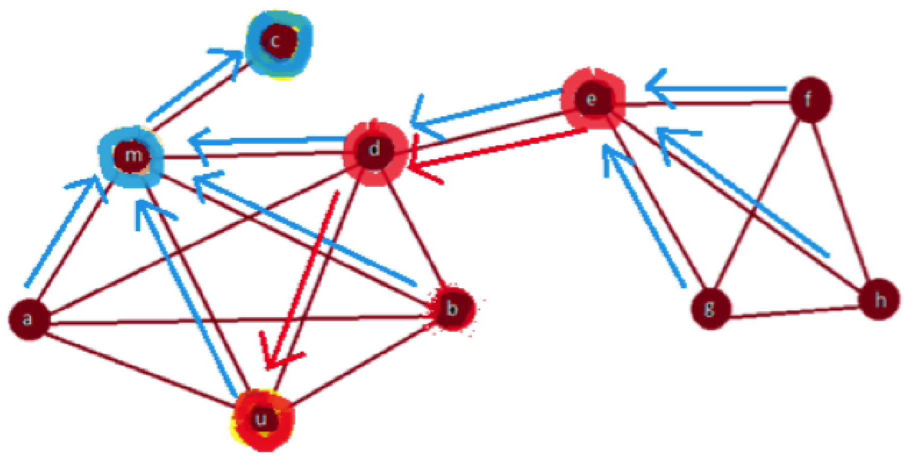


Figure 5.10: Family netowrk graph; Blue arrows represent a few of the shortest paths taken from various nodes to get to Carl, node c. Mom, node m, lies on every geodesic taken to get to node c. Red arrows demonstrates just one example [e,d,u] of all other possible node connections on whose geodesic path node m does not lie.

Conversely, Mom does not lie on the geodesic path between Ellen and you, $[e, d, u]$, so the ratio equals 0.

$$P_m(gu)/P(gu) = 0/1 = 0$$

Due to the parameters of our family network, that everyone who is immediate family has a connection, we will only find one geodesic path between all pairs. Thus, when calculating the ratio for the numerator we will only get values of 1 and 0 a node either lies on the geodesic, "1," or it does not, "0." If we had a network in which a linked pair had more than one geodesic, let say two, and our node of interest only lay on one of those shortest paths, we would find a ratio of 1/2.

As a result, if the ratio is close to 1, then node m is more important in connecting the two other nodes c and g than if the ratio is close to 0. If the ratio equals 1, then node m is essential to path between the two nodes; if the ratio equals 0, the node m is insignificant. To ratio only determine node m's importance between one pair of end nodes. In order to find m's Betweenness Centrality, we find this ratio for every possible pair of end nodes excluding m, the node of particular interest, $\frac{(n-1)(n-2)}{2}$.

Further, because each pair may have a different number of total geodesics, we must also divide each ratio by $\frac{(n-1)(n-2)}{2}$ to normalize the scale. Lastly, we sum the values from all pairs [25].

Writing out the full calculations for Betweenness Centrality can become a bit excessive since there exists (10-1)(10-2)/2=36 possible linked pair combinations for which we would have to consider. For our specific example, however, each linked pair has only one geodesic. So, Mom either does or does not lie on the linked pairs geodesic path. Therefore, we can ignore all the geodesic that exclude Mom and equate the fraction to 0. By looking at Figure 5.9, we can see that Mom is really only important in connecting other family members to grandpa. Otherwise, there exist shorts paths to other node. Our calculation thus becomes:

$$\sum_{c\neg g:m\notin c,g} \frac{\frac{P_m(cg)}{P(cg)}}{\frac{(n-1)(n-2)}{2}}$$

$$= \sum_{c\neg g:m\notin c,g} \left(\frac{\frac{P_m(ca)}{P(ca)}}{\frac{(n-1)(n-2)}{2}} + \frac{\frac{P_m(cb)}{P(cb)}}{\frac{(n-1)(n-2)}{2}} + \frac{\frac{P_m(cu)}{P(cu)}}{\frac{(n-1)(n-2)}{2}} + \frac{\frac{P_m(cd)}{P(cd)}}{\frac{(n-1)(n-2)}{2}} + \frac{\frac{P_m(ce)}{P(ce)}}{\frac{(n-1)(n-2)}{2}} + \frac{\frac{P_m(cf)}{P(cf)}}{\frac{(n-1)(n-2)}{2}} + \frac{\frac{P_m(cg)}{P(cg)}}{\frac{(n-1)(n-2)}{2}} + \frac{\frac{P_m(ch)}{P(ch)}}{\frac{(n-1)(n-2)}{2}} \right)$$

$$= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{8}{36} = .2222$$

5.4 Network Properties – Centrality Models – Katz-Bonacich and Eigenvector Centralities

Other centrality measurements such as Katz-Bonacich, Eigenvector Centrality, and as we will see later, particular forms of Diffusion Centrality all measure global centrality characteristics. Conceptually, Katz-Bonacich and Eigenvector Centrality are much more complex than those defined previously. Rather than evaluating nodes solely on their location within the network and in regards to their position to other nodes, Katz-Bonacich and Eigenvector Centralities weight different nodal characteristics and properties to identify to what extent each node impacts the network system as a whole [38]. These measures distinguish a node's centrality as ensuing from and dependent upon the centrality of its neighbors and the prominence of the arc between them [25].

What's more, these models use linear algebra to model networks and network calculations. As mentioned before, networks can be typified by a matrix where each element in the matrix represents a relationship (or lack thereof) between two nodes. If the network is constructed from direct links, then the corresponding matrix is asymmetric and elements $ab \neq ba$. If the network is made up of indirect links, however, the corresponding matrix is symmetric and elements $ab = ba$.

A matrix representation may seem less visually appealing than a standard network graph, however, matrices allow us to run centrality calculations for each node in the network at once. By calculating the centralities of each node at once, we may then compare the values to each other this comparison is the primary basis of network insight. Matrices, therefore, provide a much more quick and efficient method for network analysis. For the remainder of this paper, I will refer to g as a symmetric matrix. Mathematicians usually denote matrices by a capitalized letter. However, I chose to use g since this is the notation that Banerjee et al. (2016) use in the proofs of their Diffusion Centrality and Network Gossip models, which I will address in the later bulk of the paper.

For now, let's look at Katz-Bonacich and Eigenvector Centralities. There are two related forms of Katz-Bonacich Centrality that measure the prestige and power of a node respectively referred to in the fields to Katz Prestige and Bonacich Power. Prestige is like an elevated version of Degree Centrality and can be similarly considered as the summed influence earned for each political part in the Electoral College. The concept itself is somewhat circular. The prestige of a node is a determined by the prestige of its neighbors.

Consider, for example, the US Electoral College. Degree Centrality seen above is similar to the number of votes allocated to each state. States like Texas and California are given a large number of electoral votes because they house a greater number of people. Their larger populations show that they have a greater degree and more neighbors. For instance, although California has a smaller landmass, the state has 55 electoral votes compared to larger state Alaska, which has only 3 electoral votes [35]. In this manner, we can consider the number of electoral votes allocated to each state to be their relative Degree Centrality. If each states' electoral vote is its Degree Centrality, then the influence earned by the dominate political party of that state is like Katz Prestige. In Texas, Austin, Dallas, and Houston, the three largest cities in the state, are Democratic and tend to vote Democratic. However, because there are many more Republican cities and the sum of their influence on which way the state votes, Texas is a Republican state.

Equally, if node i has two neighbors nodes j and k, then in order to find node i's Katz Prestige, we must first know node j and node k's Katz Centralities. We normalize node j and node k's Katz Prestige values by dividing them by their respective degrees. Doing this, allows us to find the average amount of prestige they earn from each of their neighbors. These values are then summed to give us node i's Katz Prestige value.

$$P_i^K(g) = \sum_{j \neq i} g_{ij} \frac{P_i^k(g)}{d_j(g)}$$

In this equation, g represents the adjacency matrix being studied. P_i^K notes that we are looking a Katz Prestige, K, for node i. j is a place holder for all other nodes in the network and thus, $\frac{P_i^k(g)}{d_j(g)}$ looks at the Katz Prestige for all other nodes in adjacency matrix g divide by their respective degrees.

Bonacich Power is an elevated version of Decay Centrality. Instead of weighting the links in a node's geodesics, however, it weights the links in the all of walks that extend from that node (Jackson, 2008). The sum of these weighted walks determines a node's power their ability (power) to perpetuate some outcome throughout the network system. The weighting parameter is found by taking an attenuation parameter q, $0 < q < 1$, to the power of the length of the walk between node i and all other possible end nodes in the network, $(n-1)$. If the length of the walk is 1, than the weight is q^1, if the walk is 2, then the weight is q^2 and so forth. Computationally, we begin by calculating the values for all nodes a length of 1 away from the node of interest. We multiply the corresponding network matrix g by q^1 and take the dot product, $\cdot 1$, of this new matrix. When we "multiply" (i.e. take the dot product) an adjacency matrix by nx1 column vector of 1s, we are summing the column values in the matrix. Since q^1 represents the nodes that are a length of 1 from our node of interest, the value that appears in each row after taking the dot product is equivalent to $d_a(g)$, each node's degree.

Next, we sum all the nodes with a length of 2 away from our node of interest. Accordingly, when take q to the power of 2. This dilutes the parameter q by the length of the walk. We also take g to the power of 2. This captures q^2's effect on matrix g over two time periods, the amount of time required for information to travel from our node of interest to the node two links away. Then, we take the dot product of this matrix. We repeat this process for each walk length up until k, the longest walk in the network.

$$P^{K^2}(g,q) = qg \cdot 1 + q^2 g^2 \cdot 1 + ... + q^k g^k \cdot 1$$

If the network is connected then Katz's power centrality will weight a relation between node i and every other node in the network [25].

Eigenvector Centrality results from our ability to show a network in matrix form. Validation of this method extends from the Perron-Frobenius Theorem in linear algebra. It is not necessary to get into the detail of this theorem at this time, but I will provide a deeper explanation in "Eigenvector Centrality: Diffusion Centrality at T $\to \infty$".

Eigenvector Centrality applies the concept of eigenvectors and eigenvalues to network systems. Eigenvectors and eigenvalues are characteristic measurements relating to graphic modeling of multidimensional space. Think of a basic graph with y and x axis. This graph is two-dimensional (2D) and tells us something about a binary relationship. Image, however, that we are study the effects of heat on yeast and want to understand that relationship between the amount of heat applied to the expansion of yeast over time. We now have three parameters to graph, change in yeast, amount of heat, and time. Our 2D graph is no longer sufficient. Instead, we want to look at a 3D graph with y, x, and z axis.

Eigenvectors and eigenvalues tell by how much our yeast changes independent of direction when we add one more unit of heat over time. What does this mean? It means that these values tell by how much we can scale an object up or down without losing the integrity of the structure. Fundamentally, these values apply to vectors, arrows that point in the direction in space in which a object extends.

Takes this arrow for example, $\rightarrow$, and lets call it a vector. I can stretch it so that it becomes longer, $\longrightarrow$, or shrink it so that it becomes shorter. However, it remains pointing in the same direction. Eigenvalues convey by how much we stretched or shrunk the arrow. Eigenvectors are simply a single column or row matrix that captures the dependent variables that scale with the eigenvalue but do not change their directional space. For example, when we apply heat to yeast, it does not change its directional shape, but instead grows, scales up. Likewise, consider the graphics associated with the original Mario Bros video game. The graphics look 2d and quite pixilated. When Mario would grow, his graphics became more unfocused. Comparatively, the 2016 version of the Mario Bros game has scalability. The image has a 3d look and is just as clear when Mario is small as it is when he is large. This is a result of the use of vector art and graphics instead of bitmap [21] (See example in Appendix A).

With networks, eigenvectors also represent scalability; a node that is eigenvector central does not change the particular flow of information, but rather, the eigenvector central node pushes information through the network better than other nodes in the network. In other words, a node that is eigenvector central has the most expansive influence on a network like that of Malcolm Gladwell's Maven [43]. A Maven is a trusted source of knowledge, and therefore, what they say is more likely to stick with others and increase the worth of the information [13]. More importantly, these individuals are most prominently diffusively central in social networks.

Each of the previously mention centrality metrics, Degree, Bonacich Power[5]and Eigenvector centralities, are characteristic of Diffusion Centrality given different boundaries. In the following section, I will use these aforementioned centralities to develop and reinforce the definition of Diffusion Centrality. In doing so, I will also explain their roles and Diffusion Centrality's role in my adaptation of the research methodologies employed by Banerjee et al. (2016) in *Gossip: Identifying Central Individuals in a Social Network*.

[5]Referred to as Katz Centrality for the rest of the paper so as to align with notations from Banerjee et al. (2016).

Chapter 6

Model Design

”Humans socialize in the largest groups of all primates because we are the only animals with brains large enough to handle the complexities of that social arrangement.” Malcolm Gladwell,

The Tipping Point:

How Little Things Can Make a Big Difference

The hypothesis presented in Banerjee et al.(2016) is that individuals ”can identify those who are most central in a network according to ”diffusion centrality” simply by tracing gossip about people and without knowledge of the actual network structure.” The study correlates various centrality measures mathematically, identifying Diffusion Centrality as an affective and the most relevant method from which to develop their Network Gossip model. This Network Gossip model identifies those in a network who are diffusively central by tracking data from the recipients of information rather than by seeding information from an original source.

In 35 Indian villages, Banerjee et al.(2016), surveyed individuals and households with a set of questions that they developed to collect nominations for potential central individuals. They then tested the their results with the traditional Diffusion Centrality methods, by seeding information through these nominated individuals and households. By comparing these results to the Network Gossip nominations and analysis and the associated mathematics, Banerjee et al.(2006) were able to validate their new model. Their results suggest that individuals can in fact identify diffusion central nodes by tracking gossip and the number of times they hear an individual mentioned as a seed for information. As such, there does exist a relationship between Diffusion Centrality and Network Gossip, and overall Diffusion Centrality accounts for some of the network effect of information sourced by nominated individuals and households from the Network Gossip model.

6.1 What is Diffusion Centrality

Diffusion models track the dispersion of information, diseases, trends, etc. about networks over time. When studying the dispersion of information specifically, diffusion models looks at how information known initially to a single or handful of individuals is transmitted and propagates throughout a network over time when each person in the network has a choice to spread or withhold information from others [1]. Diffusion Centrality attempts to distinguish individuals in the network who most influentially spread information to others.

We can image that if the network of interest were a cycle then the spread of information throughout the network would be similar to a game of telephonethe source of the information would pass it along to either one or both of the people sitting next to him/her, and then they would share the information with the person sitting next to them and so forth until everyone has heard the information. What is important and different about Diffusion Centrality (its relation to the Network Gossip model) and other SNA models is that they pick up nuances of choice behavior that arises when studying real-life networks, unconstrained by the rules of the game of telephone.

Traditionally, in the game of telephone players sit in a circle and can communicate only to those on either side of them, a cycle network. Rule dictate however, that each person may only listen from one of the two people sitting next to him/her and can only talk with the other. As a result, there is a clear path in which information must flow in order to reach everyone in the circle. If players decided on a clockwise flow of information, that is each person heard the telephone message from the person on their right and shared it with the person on their left. If this cycle network represented a real life circumstance rather than a game, however, the source of information could choice whether he/she wanted to share the message with person on his/her left or right. Those two individuals, if after the first decision by the source heard the message, then they two have a choice to pass along the information to the individual on either side of them. They can share the information with the person sitting farther away from the source who has not yet heard the message, or they could share the information with the source. Since the source of information already knew the message, by sharing it with him/her creates an "echo." Theses choices to share forward, share backwards, or withhold information continue until everybody in the network has heard the information, or at least until choices made by the individuals involved prevent the information from spreading about the network any further. For example, for some reason the source decided he did not want to talk to the either person on his/her left or right, then no one in the circle would ever hear the message and the network would consequently not exist.

This example provides a relatively easy image for understanding choice behavior since individuals in the cycle network are only connected to two other people. In reality, individuals have many more connections and links, maybe even hundreds depending on the context in which you look. As such, network structures that are also not perfectly circular and it can be much harder to identify the distribution and directions in which the information will travel, and is also why these seemingly difficult diffusion models actually make network comprehension and analysis much more doable and digestible.

The diffusion model we use in our research is Diffusion Centrality, a simplified version of Banerjee et al.(2016)'s Communication Centrality model. The Diffusion Centrality model simply predicts information spread based on the probability that two individuals are connected and the probability that they communicate with each other. Communication Centrality and other more in-depth models, however, take into account a variety of other factors such the ability for nodes to learn and behave responsively. The inclusion of choice behavior allows for the quantification and simulation of previously abstract social studies. The in-depth models provide for real world application such as in marketing; marketing fields, for instance, use complex diffusion models to understand not only how product information circulates about a network via word of mouth, but also how likely markets and individuals are to adopt and purchase products [9]. Banerjee et al.(2013)'s Communication Centrality model, however, is difficult to reproduce without parameters that exactly parallel those in their study. Instead, Banerjee et al.(2016) enhanced their general Diffusion Centrality model so that it too incorporates choice behavior in a model they call Network Gossip. In the following sections, I will explicate, mathematically, why Banerjee et al.(2016) uses Diffusion Centrality as the foundation for their Network Gossip model, and why Network Gossip is the most scrupulous method for my study around communication about environmental sustainability at LMU.

6.2 Diffusion Centrality as a Foundation for Identifying Central Individuals Through Network "Hearsay"

The following is a list of general abbreviations and their alternative forms that I will use throughout the remainder of the paper:

H Hearing Matrix

DC Diffusion Centrality

$d(g)$ Degree Centrality for adjacency matrix g

KB Katz-Bonacich Centrality

$v^{(1)}(g)$ Eigenvector Centrality for adjacency matrix g

- $v^{(1)}$ first right hand eigenvector that corresponds to λ_1, and the greatest effect in a symmetric, diagonalizable matrix according to the Perron-Frobenius Theorem (detailed in Appendix L); Eigenvector Centrality $\approx$ Diffusion Centrality as $T \to \infty$); The value of the greatest amount of variance across the network data; $v^{(R,1)}$.
- $v^{(2)}$ first left hand eigenvector; $v^{(L,1)}$

p discrete probability that two individuals within the network are linked q probability that linked individuals communicate T # of periods t time

The diffusion model that I adopt and adapt to analyze the ES communication network at LMU is the Network Gossip model which Banerjee et al.(2016) derive from their Diffusion Centrality model, which in turn is derived from their Communication Centrality model devised in their study *The Diffusion of Microfinance* (2013). Diffusion Centrality is a proxy for Banerjee et al.(2013)s communication centrality measurements. The model strongly correlates to Communication Centrality and maintains much of Communication Centralitys predictive capabilities, but, due to reduced parameter constraints, is adaptable to various samples and forms of data collection. However, to capture choice behavior forgone by slackened parameters in the Diffusion Centrality model, I chose to use Banerjee et al.(2016)s Network Gossip model. The Network Gossip model will allow me to uncover if there are differences between the formal and informal communication network structures about LMUs environmental sustainability. An understanding of the derivation of the model from Diffusion Centrality and the models correlation will further support this research choice for my study.

Network Gossip essentially builds and identifies central individuals in reverse of Diffusion Centrality. Methodologies typically associated with Diffusion Centrality start by seeding information to various individuals within a network and following the spread of that information from individual to individual across the network. Data collected by tracking the dispersion of the information resulting from each individual after a determined time period provides the necessary information to run Diffusion Centrality calculations and to determine which individuals in the network may be the most diffusively central. Network Gossip, on the other hand, begins by asking individuals within a network a variety of questions to which they identify, nominate, others as network mavens. Banerjee et al. (2016) verify that these nominated individuals in fact are the most diffusively central within the network at $T \rightarrow \infty$ and not nominated simply as a result of having the most number of friends or because of leadership roles within a community.

What is important to note, is that this holds true only at $T \rightarrow \infty$.This makes sense due to the nature of the data collection method. The individuals surveyed and asked to nominate others within the network have a probability, p, of being a part of the network themselves and then a probability, q, of communicating with individuals within that network. The probability that each individual is linked in the network, p, is independent as modeled by Erdos-Renyi[1]. The probability that each linked persons i,j communicate, q, is distinct and not determined by individual i and individual js other connections, but instead by the length the edge between them.

Over time, as the information begins to flow through the network and away from the original source, the probability nodes continue to pass and communicate information continues at a diminishing rate similar to the weighted walks used in Katz Centrality (whose mathematical depiction and correlation we will address later). However, although each probability at time period T+1 is smaller than the probability at previous time period T, the sum of the probabilities from T=1 until the current time period is greater. Accordingly, while it is less likely that nodes communicate as time increases, it is more likely that nodes in the network have already communicated with and have heard the information. In other words, it is more likely for an individual to be a part of a network after a longer period of time. Thus, because we are currently unable to determine during which time period our surveyed individual heard the information originating from their nominee, it is better to consider the survey responses and data from our Network Gossip model for a time period in which the connection between two nodes has a higher probability of being realized.

If we were to correlate Network Gossip to Diffusion Centrality at $T = 1$, we would lose the ability to confirm that an individual is diffusively central to the network as a

[1]Paul Erdos and Alfred Renyi introduced their model for graph theory in 1959. It posits that all possible graphic structures of a network are equally likely when there is a fixed number of nodes and a fixed number of links (Chapter 1 Overview, 2016). Edgar Gilbert added to their model later on, suggesting that each possible link has an equal and fixed probability of being present or absent in the network [31]. That is, when some parameter restricts outside influence on the network so that no node made leave or entry the network during the time studied, the number of nodes present in the network remain fixed. With a fixed number of nodes within the network and no outside influence affecting the network, the maximum number of links possible in a connected network is also fixed. However, it is not certain that the network will be connected. Instead, the network structure can take many different forms depending on which nodes are linked together. The Erdos-Renyi model suggests that the probability of linking any two nodes in the network (when the number of nodes and possible links are fixed) and thus generating any one of the various network structures is equal [3].The model is represented by $G(n, p)$, denoting all possible network graph with n nodes and p probability of a link between each pair of nodes independent of the other links. When n and M, the number of links, are fixed, the probability of generating each graph is: $P^M(1-p)^{(n-2)-M}$

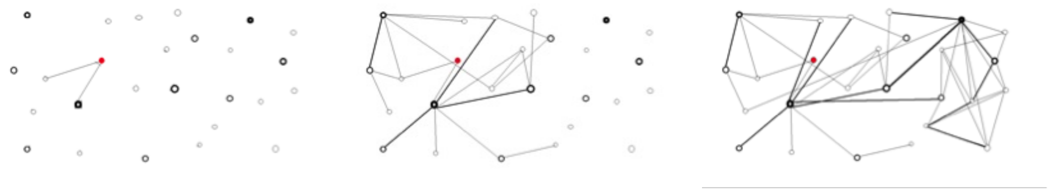


Figure 6.1: T=1, $1 < T < \infty$, $T \to \infty$

whole. Since individuals only speak with others to whom they have a direct link, at $T = 1$, we could only examine Diffusion Centrality as it amounts to Degree Centrality.

$$DC(g; p, 1) = pd(g); T = 1$$

At T=1, a nominee could have only communicated with others with whom they are linked. Consequently, if we were to assume T=1, a person surveyed through our Network Gossip model would be an individual directly linked to our nominee, and thus the nominees centrality could only equate to that of Degree Centrality.

$$d_i(g) = \sum_j g_{ij}$$

At the other extreme, we could look at the Network Model at $T = \infty$. However, for practical reasons, this is not possible. We cannot collect data at $T = \infty$ simply because we can never reach $T = \infty$, and therefore, can never address this case beyond theoretical, mathematical observation. Banerjee et al (2016) do convey, however, that if q¡$\frac{1}{\lambda_1}$, Diffusion Centrality at $T = \infty$ is proportional to Katz Centrality described previously.

$$DC(g; q, \infty) = KB(g, q); q < \frac{1}{\lambda_1}$$

The correlation above suggests that a nominee in our Network Gossip model at $T = \infty$ will be not just diffusively central, but also exemplify characteristics of Katz Centrality, a prestigious node who sustains the lifetime of a piece of information within the network. I will show and explain Banerjee et al.(2016)s Katz Centrality equation, but must admit that it is much more digestible to understand conceptually, and so, I will attempt to illustrate the concept first.

Think of the real life, no rules game of telephone example at the beginning of Diffusion Centrality. When a player choice to share the information backwards, repeating the message to the person who had first told him/her the information, the network experiences what network scientists call echoes. If someone in the network continuously chooses to knowingly or unknowingly share information with others who have already heard the message a first time, this person persists in keeping the information active; if enough time passes so that every other individual has heard the information, or at least until choices prevent the information from spreading about the network any further, and this individual continues to repeat the information to others then he/she is Katz Central. In other words, the information stays relevant and top-of-mind so long as this Katz Central individual keeps repeating and sharing old news. Banerjee et al. (2016) defines this mathematically since realistically we cannot collect data at $T = \infty$, the time period at which it is certain that information has had enough time to saturate the network and at which network echoes begin to occur.

At $T = \infty$, the network has enough time to realize the full capability of information movement around the network by the individuals involved. That is when $T = \infty$, we can see the network effect in a holistic sense: all those who would communicate have with a high probability already communicated, all those who would hear the information at least one time if not more have with a high probability already heard the information and so on. When we set $q < \frac{1}{\lambda_1}$, we are noting the aforementioned: the probability that information diffusion does occur after$T = \infty$ must be less than the greatest effect on the network, λ_1, as this has already been realized. Katz Centrality is thus an enhancement on Diffusion Centrality and not an essential measurement in identifying individuals who are diffusively central. It is a sufficient, but not necessary condition for Diffusion Centrality. Therefore, it is not required that we confirm the correlation between Diffusion Centrality and Katz Centrality at $T = \infty$ and $q < \frac{1}{\lambda_1}$ in application beyond that which is theoretical and mathematical for our Network Gossip Model.

Individuals who are Katz Central when computing

$$KB(g,q) := (\sum_{t=1}^{\infty}(qg)^t) \cdot 1$$

after network saturation, $T = \infty$, are key to keeping alive information within networks. Essentially, these individuals do not let information become peripheral or a thing of the past, but instead has the power to create information echoes and maintain information relevance within the network.

6.3 Eigenvector Centrality: Diffusion Centrality T to Infinity

The tipping point is that magic moment when an idea, trend, or social behavior crosses a threshold, tips, and spreads like wildfire. Malcolm Gladwell, The Tipping Point: How Little Things Can Make a Big Difference

Since we have ruled out $T = 1$ and $T = \infty$ as appropriate periods of time in which to determine Diffusion Centrality from our Network Gossip model, we now turn to looking at $T \to \infty$ as an appropriate time frame for our analysis. We will see further that this is in fact the most appropriate time period in which to assess our Network Gossip model since the network has enough time to realize patterns and tangible data is accessible. Furthermore, at $T \to \infty$, Diffusion Centrality and Katz Centrality converge with Eigenvector Centrality, a relatively manageable model.

Eigenvectors identify the extent to which a matrix and the elements within that matrix can stretch without changing direction. According to Oscar Perron and George Frobenius, any real square matrix will have a unique largest real eigenvalue with a corresponding eigenvector, which can be chosen to have only positive components [25]. We want to prove the claim that there exists a unique largest real eigenvalue and corresponding eigenvector because, if such a property holds, it will allow us to determine the greatest extent to which each individual in our network influences the spread of information throughout the network. Accordingly, proof of the Perron-Frobenius Theorem is as follows. If the network of interest forms a square nonnegative stochastic matrix, we begin the proof by looking at the matrixs associated characteristic polynomial and roots. Suyeon Khim from the University of Chicago exemplifies this proof by starting with matrix B,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

and characteristic polynomial, $p_B(\lambda) = det(\lambda I - B) = \lambda^2 - (a+d)\lambda + (ad+bc)$.By factoring the characteristic polynomial using the quadratic formula, we get two roots for $p_B(\lambda)$:

$$\lambda(B) = \frac{(a+d) + \sqrt{b^2 + 4ac}}{2} \quad \& \quad \lambda^{'}(B) = \frac{(a+b) - \sqrt{b^2 + 4ac}}{2}$$

Since B is nonnegative and thus only has positive entries, $\sqrt{b^2 + 4ac}$ is positive, confirming that values for λ are real. The proof continues to verify the claim that there exists a corresponding eigenvector with strictly positive entries (See Appendix B)

We can test the existence of a steady state when applying the Perron-Forbenius Theorem to Markov chains. A Markov matrix is nonnegative and column stochastic where each column sums to 1. A Markov chain occurs when a vector is transformed by an unchanging transition matrix at each time period. Eventually, the vector itself will at some point become unchanging itself and reach a steady state. At each period, the matrix columns sum to 1 and at the steady state the eigenvector sums to 1 [27]. The transformation that accomplishes this is similar to taking a matrix to the power t as we do in our centrality models where algebraically we cannot distribute the power value, but instead must multiply the matrix by itself t number of times.

Accordingly, Banerjee et al.(2016) and I use matrix g as our transition matrix. $v^{(1)}$ denotes the first right hand eigenvector for the matrix. λ_1 is the largest eigenvalue that specifically identifies the value of the largest stretch on the matrix where each λ is essentially a proportionality factor that identifies by how much an individuals network stretch influence is determined by his/her neighbors network stretch influence [4].

More importantly, this trait, the identification of a largest, positive eigenvalue, allows us to set the probability that individuals communicate, q, relative to $\frac{1}{\lambda_1}$ in our different centrality models. When q=1, we are expressing that information dispersion will equal that of the largest effect in the network. Recall again that we have already determined that any effect beyond λ_1 occurs at $T = \infty$. Furthermore, should we delineate Katz Centrality with the parameters $q \geq \frac{1}{\lambda_1}$ rather than $q < \frac{1}{\lambda_1}$ and $T \rightarrow \infty$ rather than $T = \infty$, we see that it converts to Eigenvector Centrality.

By observing λ_1 at $T \rightarrow \infty$ we can really begin to perceive network patterns, characteristics, and traits, and in terms of our Network Gossip model, it is during this time period that we can endorse our participants survey data with the greatest assurance because there would pass enough time that the probability of each individual connecting and communicating would be sufficiently large. To support this concept, Banerjee et al.(2016) again cite Erdos and Renyi[2]. The theorem suggests, the average distance between most nodes is almost the same as the diameter. We use the expected value of the diameter since the true value of the diameter is not known with certainty in our backwards Network Gossip model. By setting $q = 1/E[\lambda_1]$ and

$T = E[Diam(g)]$ as benchmark parameters, we can ascertain that if $T < E[Diam(g)]$, there would not have passed enough time for us to observe meaningful information flow beyond that related to Degree Centrality. When $T = E[Diam(g)]$, T is the number of periods in which just enough time passes for each individual in the network to have had the opportunity to talk to all other individuals in the network [2].

As $E[Diam(g)] \to \infty < T$, the rate of information diffusion slows, the network approaches complete saturation and the spread of information must double back into the network. At this point, the individuals hear echoes. In other words, individuals who hear the information have already once before heard the same information [2]. It is also at this time that we can begin to look for λ_1, and again, we use $q \geq \frac{1}{\lambda_1}$, where λ_1 is bounded by the networks diameter and is thus $\lambda_1 E[Diam(g)]$. When we set $q \geq \frac{1}{\lambda_1}$, we define the probability that individuals in the network communicate to be greater than the inverse of the greatest effect on the network. Meaning that, as $T \to \infty$ there exists a greater probability that individuals will share and spread the information about the network and do so at a higher velocity since the largest network effect has yet to be realized.

Now that we have covered the logic as to why it is best to back into Diffusion Centrality from our Network Gossip data through Eigenvector Centrality, lets examine

[2]In order to complete this proof, we must first understand Poisson distributions. Poisson distributions and it corresponding regression model is of upmost importance in the regressions of the study, and therefore, I go into much more detail in that chapter. What is important here, Poisson distributions are random. That is the placement/value of one node is not related to or dependent upon the placement/value of another node.

The Erdos-Reyni Theorem suggests that the average distance between nodes in the network is equivalent to the network diameter. In other words, if we know the network diameter, then we know the average distance between most nodes in the network. So, we seek to find the average distance between nodes in the network. We let D denote the degree of a node from the Poisson distribution. Since a link either exists or does not exist with a probability p and $1 - p$ respectively, the expected degree, $E[D]$, of a single node $= (n-1)p$. In order to find the average degree, we need to run the expected value for all nodes in the network. Consequently, we can predict the average degree, and thus, the network diameter, by using a simple probability combination equation across all possible degree values in the network and p and $(1-p)$ as the respective probabilities:

$$P(D = d) = \genfrac{}{}{0pt}{}{n-1}{d} p^d (1-p)^{n-1-d}[3].$$

the correlation mathematically. Notice the equation above. It puts the aforementioned concept into an arithmetical, computable notation, and is proof that what was said in words works mathematically as well. It is clear that the right hand side of the equation is the first, right hand vector (associated with matrix g from our model). So, let us begin by breaking down the left hand side of the equation, which correlates Diffusion Centrality to Eigenvector Centrality, and which is best done by defining the elements shown in the middle section of the equation

Recall that in the Diffusion Centrality model, g is an adjacency matrix, and as such carries the following properties: it is positive, symmetrical, and diagonalizable.

$$g = V \wedge V^1$$

V is a matrix of gs column vectors, $\wedge$ is the diagonalized matrix of g where the associated eigenvalues are the diagonal elements and all other elements are zero, and V^{-1} is the inverse of V. This representation suggests that matrix g is similar to its diagonal matrix, a much simpler matrix to use throughout our research.Accordingly, the Diffusion Centrality matrix calculated using g, DC(g;q,T), is also positive, symmetrical, and diagonalizable, and therefore, so is the numerator. In the denominator, $\frac{q\lambda_1-(q\lambda_1)^{(T+1)}}{1-q\lambda_1}$ defines the change in the probability of the network realizing the largest effect, $q\lambda_1$, from each time period divided by the probability that the largest effect is not realized at the current time period. As noted before, expected probability that individuals in the network communicate as $T \to \infty$ is greater than the inverse of the largest affect on the network, $q \geq \frac{1}{\lambda_1}$. In other words, $q\lambda_1 \geq 1$ where $q\lambda_1$ is increasing (at a decreasing rate) for each T as $T \to \infty$ before λ_1 is realized. This value increases at a decreasing rate as the network becomes more and more saturated and there are fewer unaffected individuals and links. Since $q\lambda_1$ is therefore larger with each addition

$T + 1$, $q\lambda_1 - (q\lambda_1)^{(T+1)}$ is negative; and, $1 - q\lambda_1$ is also negative, making the entire denominator expression positive.

By dividing the Diffusion Centrality vector by the percent change in the network λ_1 effect, we may identify by how much each individual in the network contributes to the spread of information at different time periods[3]. By computing the equation at $T \to \infty$, we generate the value for the full extent of each individuals influence in the network given that $T \to \infty$ is sufficient time for each individual to achieve his/her full fitness; and is thus the definition of our first right hand vector.

In Figure 6.4, the blue S Curve shows this effect. At the beginning, when only a few nodes know and are sharing the information, diffusion is relatively slow. However, as more nodes hear the information and more untouched nodes become accessible during each time period, the speed at which the information permeates the network can increase exponentially. This is the tipping point. As $T \to E[Diam(g)]$, considered previously, and only a few nodes is the network remain uniformed of the information, diffusion of the information begins to slow and the number of informed increases and a decreasing rate until the network is completely saturated. The black Bell Curve shows the respective rate at which individuals in the network hear the information. The early adaptors are those who first hear and know the information. Once we hit the tipping point, and the rate of diffusion increases, a larger number of individuals in the network

[3]The properties explicated the discussion of the largest network effect

$$\lim_{T\to\infty} \frac{DC(g;q,T)}{\sum_{t=1}^{T}(q\lambda_1)^t} = \lim_{T\to\infty} \frac{DC(g;q,T)}{\frac{q\lambda_1-(q\lambda_1)^{T+1}}{1-q\lambda_1}} = v^{(R,1)}$$

are corollary to Erds-Renyis Theorem. In this study, we focus on a probability that individuals communicate that is greater than the inverse of the largest network effect, $q \geq \frac{1}{\lambda_1}$. Accordingly, the expected average Diffusion Centrality may tend towards infinity: if $\frac{1}{E(\lambda_1)} = o(q)$, then $E[DC(g(n,p);q,T)] \to \infty$ where $E(\lambda_1) = np$ (n number of nodes times the probability that nodes within the network are linked). Interesting, however, if $q = o(\frac{1}{E(\lambda_1)})$, then $E[DC(g(n,p);q,T)] \to 0$. $\frac{q\lambda_1-(q\lambda_1)^{T+1}}{1-q\lambda_1} \to 0$ if $q\lambda_1 \to 0$, suggesting that there does not exists the possibility of the network realizing the largest network effect (perhaps due to weak communication between nodes). Consequently, the denominator expression is negative, causing $\lim_{T\to\infty} \frac{DC(g;q,T)}{\frac{q\lambda_1-(q\lambda_1)^{T+1}}{1-(q\lambda_1)}}$ to be negative and $E[DC(g(n,p);q,T)] \to 0$ [2]. Thus, this expression importantly exemplifies that when there is a failure in communication between nodes information diffusion will likely not occur.

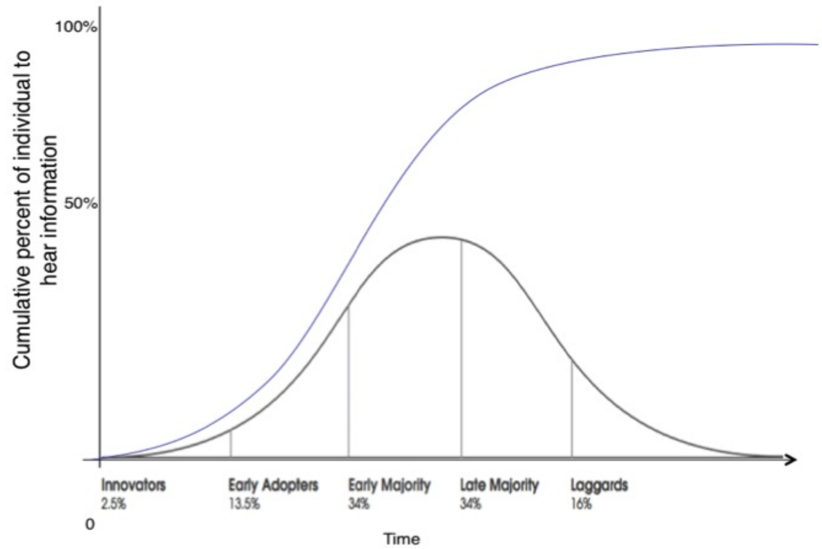


Figure 6.2: Diffusion S curve graph; The Y-Axis is the cumulative percentage of nodes who have heard the information, and the X-Axis is time over which the information spreads and stays relevant within the network.

hear the information and thus, we get the early majority. As information spread continues and rate slows, a large portion of the remaining uninformed nodes hear the information, the late majority. And finally, as diffusion rate reaches a terminal rate, the laggards, the remaining nodes in the network will hear the information if at all.

6.4 Building a Hearing Matrix from the Network Gossip Model

So far, we have unraveled the different centrality methods that connect Diffusion Centrality to Network Gossip under the intention to identify if we can accurately correlate our two models. However, it is not enough to say that because we have confirmed that Diffusion Centrality tends towards Eigenvector Centrality at $T \to \infty$ and that Eigenvector Centrality most accurately translates the number of nomination an individual might receive through our Network Gossip model to equal their respective effect though Diffusion Centrality modeling. As I have mentioned throughout the previous section, the key difference between Diffusion Centrality and Network Gossip is the direction from which the methods approach their analysis, following the information as it spreads outward from the original source or by tracking the information backwards from network nodes who hear it. In order to truly confirm the validity of the Network Gossip model, we must also explicate the relationship mathematically.

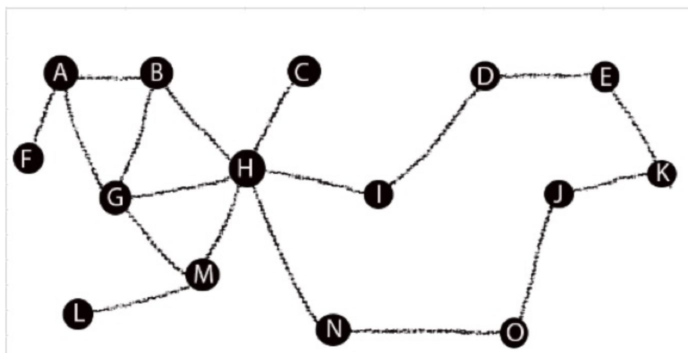


Figure 6.3: Example network graph.

Let's use Figure 6.5 as an example in the following explanation. When we build an adjacency matrix based on individuals' links within the network, multiply it by the probability that the linked individuals communicate, and track the effects on the network of periods of time by taking the matrix to the power t, we are able to predict in what way different individuals in the network spread information. This however, does not capture the reverse, the predictive level to which an individual in the network hears information originating from another individual in the network. For this reason, Banerjee et al.(2016) established what they call the Hearing Matrix. The Hearing Matrix, H, is a 1xn column matrix, calculated from the original adjacency matrix. It tracks "the expected number of times [node] j hears a piece of information originating from [node] i" [2].

$$H(g; q, T) := \sum_{t=1}^{T} (qg)^t$$

Like in the Diffusion Centrality model, g is the adjacency matrix for our network and carries the same positive, symmetrical, diagonalizable properties.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0
B	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
D	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
E	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
F	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
G	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0
H	0	1	1	0	0	0	1	0	1	0	0	0	1	1	0
I	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
K	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
M	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
N	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1
O	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0

Figure 6.4: Matrix g associated with the example network graph above.

Similarly, q is the probability that two nodes in the network communicate since communication is still necessary for an individual to then hear and acknowledge a new piece of information. The assembly of this matrix makes logical sense by our Network Gossip model in which individuals can only nominate others in the network if they are aware that they have heard information originating from that source. However, in order to equate our Network Gossip model to the Hearing Matrix requires a bit of further specification. When using the Network Gossip model, survey participants may nominate more than one individual, but because these nominations remain distinct, we can thus recognize each nomination as different links in which there can only exist the survey participant and the nominated individual as the outermost nodes in the link.

As such, we can use the network gossip data from each survey participant to build our Hearing Matrix where the network gossip data from node A, $NG(g; q, T)_A$, forms the Ath column of the matrix.

$$NG(g; q, T)_{\cdot A} = H(g; q, T)_{\cdot A}$$

If in the past, node A has heard information that originated from node H as well as a piece of information that originated from individual N, node A is likely to nominate both H and N in our survey. For our Network Gossip model, we construct a network graph that represents an undirected link between A and H and A and N. The hearing matrix is simply the network matrix for this model; it shows a 1 in cells AH and HA, and AN and NA for the undirected relationship between A and H and A and N. Once we build our matrix, including data for every individual in the network, we can run the previously mentioned centrality measurements to find out from whom individual A is most likely to hear information originating from for each given time period.

Figure 6.6 shows an example matrix g at T=1, Degree Centrality. Looking at column A and rows H and N, we see that at the 1 represents their undirected link. If we want to figure out how likely A is to hear from H and N after some time, i.e. H and N's Eigenvector centrality (Diffusion Centrality at $T \to \infty$), we compute $H = \sum_{t=1}^{T}(qg)^t$ using MatLab with q = $\frac{1}{E(\lambda_1)}$ and $T = E[Diam(g)]$, Banerjee et al.(2016)'s benchmark parameter. For matrix g, q = 1/3.143 and T = 7. $H(g; q, 7)_{HA} = 1.0080$ and $NG(g; q, 7)_{NA}$= .2860. (See Appendix C for full matrix g at T = 7) Accordingly, node A is 3.5 times as likely to hear information originating from node H compared to individual N.

$$1.0080/.2860 = 3.5245$$

When we take the dot product of matrix g by a vector of ones, we find that node H and node G have the highest values, 10.9082 and 10.0556, respectively. (See Appendix D for full list of values and codes). Therefore, it more likely that at T = 7 most other nodes in the network have heard information originating from node H or node G over all other nodes.

6.5 Adjoining Diffusion Centrality to Network Gossip Through the Construction of a Hearing Matrix

We ran this model with the benchmark parameters suggested by Banergee et al.(2016), $q = \frac{1}{E(\lambda_1)}$ and $T = E[Diam(g)]$, but we still need to test the uniformity between the Network Gossip and Hearing matrix over time. Earlier, we established that it only makes sense to analyze our Network Gossip survey data collected at $T \to \infty$. Eigenvector Centrality best discerns network central individuals. Thus, we must also check whether the hearing matrix also correlates with Eigenvector Centrality and converges to λ_1 at $T \to \infty$ mathematically.

We once again use the properties of our matrix g, $g = V \wedge V$, to help verify the relationship between the Hearing matrix and Eigenvector Centrality. For simplicity, we will also equate $q\widetilde{\lambda_k} = \widetilde{\lambda_k}$ where $q\widetilde{\lambda}$ is the probability that individual k will hear information to his/her greatest extent possible within the network. Accordingly, we set

$$H = \sum_{t=1}^{T} (qg)^t = \sum_{t=1}^{T} (\sum_{k=1}^{n} v_i^{(R,k)} v_j^{(L,k)} \widetilde{\lambda}_k^t)$$

.

The right side of the equation above is simply an expansion of

$$\sum_{t=1}^{T} (qg)^t$$

. The outside notation is the same, the sum of the function of time periods T beginning with t=1. Inside the Sigma, we deconstruct our matrix g into its associated right hand and left hand vectors. When we multiply the right hand vector by the left hand vector, we get a 1x1 matrix, a single value. This value is the expected number of times that individual k will hear a piece of information from others within the network. By multiplying this value by $\widetilde{\lambda_k}$, we finding the probability this network effect occursthat individual k will hear the information this many times. Since the Perron-Frobenius Theorem confirmed that matrices, such as our matrix g has a largest unique eigenvalue, we can further expand the expression above so that

$$H_{\cdot j} = \sum [v^{(R,1)} v_j^{(L,1)} \widetilde{\lambda}_1^t + v^{(R,2)} v_j^{(L,2)} \widetilde{\lambda}_2^t + O(|\widetilde{\lambda}_2|^t)]$$

and where the eigenvalues are ordered from largest to smallest. Since we have ordered the eigenvalues from largest to smallest, the first part of the expression is trivial; it is the largest eigenvalue for matrix g. The second part of the expression is simply notating the second largest eigenvalue for our matrix g. Lastly, the third part of the expression suggests that because, as we move further along, our eigenvalues get smaller and smaller their values approaches a lower bound and their effects become negligible.

We can rearrange the expression even further by pulling known values out in front of the Sigma and dividing by the largest eigenvalue.

$$= \frac{v^{(R,1)} v_j^{(L,1)} \sum_{t=1}^{T} \widetilde{\lambda}_2^t}{\sum_{t=1}^{T} \widetilde{\lambda}_2^t} + O\left(\frac{\sum_{t=1}^{T} |\widetilde{\lambda}_2|^t)}{\sum_{t=1}^{T} \widetilde{\lambda}_1^t}\right)$$

$$= \frac{H_{\cdot,j}}{\sum_{t=1}^{T} \widetilde{\lambda_1^T}} \to v^{(R,1)} v_j^{(L,1)}$$

Consequently, we find that individual j's number of time of information heard from any another individual in the network divided by the probability of individual j attaining the full scope of their hearing ability closely equates to the production of its right hand and left hand eigenvectors. $O(\frac{\sum_{t=1}^{T}|\widetilde{\lambda_2}|^t}{\sum_{t=1}^{T}\widetilde{\lambda_1^t}})$ drops from the equation because $\widetilde{\lambda_1} > 1$ and $\widetilde{\lambda_2} > \widetilde{\lambda_2}$, and thus approaches 0. At this point, we can justify the parallel between Network Gossip and Diffusion Centrality. Recall that Diffusion Centrality for an individual i "is the expected number of times that a piece of information that originates from i is heard received by all other individuals in the network" at given time period [25]. If we sum the columns in the hearing matrix by multiplying the matrix by a column vector of 1s, we can find the total number of times that information from individual i was heard by the other individuals in the network.

$$H(g;q,T)\cdot 1 := (\sum_{t=1}^{T}(qg)^t)\cdot 1$$

And, thereupon, we can postulate that each individual's summed value in the column vector is his/her expected diffusion centrality value.

$$DC(g;q,T) := H(g;q,T)\cdot 1 = (\sum_{t=1}^{T})\cdot 1 = NG(g;q,T)$$

It is through this notation that we present Banerjee et al.(2016)'s original hypothesis that individuals "can identify those who are most central in a network according to "diffusion centrality" simply by tracing gossip about people and without knowledge of the actual network structure" [2]. Still, we must validate this relationship to see if it holds beyond notation. One such way to do so is by testing the covariance between our Degree Centrality and Network Gossip models.

6.6 Testing the Relationship Between Diffusion Centrality and Network Gossip

We have shown that we can relate Diffusion Centrality and Network Gossip by rearranging their properties. However, we still need to check if the proposed relationship holds mathematically. We do this by testing covariance, Cov(DC(g;q,T), NG(g;q,T)), where we multiply together the variances of each model.

$$Var(DC) = DC_i - \sum_k \frac{DC_k}{n}$$

and

$$Var(H) = H_{ij} - \sum_k \frac{H_{kj}}{n}$$

And where $\sum_k$ states that we find the sum of Diffusion Centrality and Network Gossip divided by the total number of nodes for each node in the network so that $\sum_k \frac{DC_k}{n}$ and $\sum_k \frac{H_{kj}}{n}$ are equivalent to mean for each model.

If Diffusion Centrality and Network Gossip are not at all related we can expect to get a value of 0 for covariance, which thus says that any variance or discrepancy in the predicted values of one model cannot be explained by those in the other model. If we get back a negative value for the covariance between Diffusion Centrality and Network Gossip, then the two models are related but inversely so. If the predicted value in one model varies positively, then we can expect the value in the other model to change in a negative direction. What we hope to find is a positive covariance value, which would suggest that disparities found in the models predicted values can be explained by those found in the other model, and we would be able to say more definitively that our Diffusion Centrality and Network Gossip models are reciprocal.

We test Cov(DC(g;q,T), NG(g;q,T)) by comparing the models for a single individual in the network, for example the Diffusion Centrality of individual i and the hearing of information from i by individual j, where we postulate the number of times that individual j hears information originating from individual i is a homologous to individual is Diffusion Centrality.

$$DC_i = \sum_j H_{ij}$$

Banerjee et al.(2016) suggest, however, that it is best to test covariance between Diffusion Centrality for individual i with the sum network gossip measure rather than that directly derived from the link between individuals i and j because of Diffusion Centralitys sensitivity to different analysis at different time periods. Recall the problem we solved when measuring Diffusion Centrality at $T \to \infty$, correlating the measurement to Eigenvector Centrality in the previous section. It is for this same reason, that we use the summed Network Gossip measure: at any specific time period before $T \to \infty$, we cannot assume that the network would have had enough time to reach network saturation and therefore, not all individuals in the network would have had the opportunity to have heard the information, and the full scope of is Diffusion Centrality will be realized.

Subsequently, we use the forms,

$$DC = (\sum_{t=1}^{T} (qg)^t) \cdot 1$$

and

$$H = \sum_{t=1}^{T} (qg)^t \sum_{t=1}^{T} (qg)^t$$

to find the respective covariance, finding that there exists a positive value and thus, a positive relationship between Diffusion Centrality and Network Gossip. By testing the covariance between the Diffusion Centrality and Hearing Matrix, we found that there does exist a positive correlation between the two models. Furthermore, since there exists a positive relationship between Diffusion Centrality and the Hearing Matrix when correlated with Eigenvector Centrality at $T \to \infty$, and the Hearing Matrix is representative of Network Gossip when correlated with Eigenvector Centrality, we can justifiably affirm that there also exists a positive relationship between Diffusion Centrality and our Network Gossip Model[4]. And, more importantly, that the Network Gossip model is a good estimate of an individuals Diffusion Centrality within a network.

[4]The proof, solved by Banerjee et al. (2016), is as follows:

$$DC = (\sum_{t=1}^{T} (qg)^t) \cdot 1$$

and

$$H = \sum_{t=1}^{T} (qg)^t$$

$$cov(DC, H_{\cdot j}) = \sum_{j} (DC_i - \sum_{k} \frac{DC_k}{n})(H_{ij} - \sum_{k} \frac{H_{kj}}{n})$$

$$\sum_{j} cov(DC, H_{\cdot j} = \sum_{i} (DC_i - \sum_{k} \frac{DC_k}{n})(\sum_{j} H_{ij} - \sum_{k} \frac{\sum_j H_{ij}}{n})$$

$$\sum cov(DC, H_{\cdot j} = \sum_{i} (DC_i - \sum_{k} \frac{DC_k}{n})(DC_i - \sum_{k} \frac{DC_k}{n}) = var(DC)$$

Chapter 7

Regressions

7.1 Authenticating the Relationship Between Network Gossip and Diffusion Centrality

"The tipping point is that magic moment when an idea, trend, or social behavior crosses a threshold, tips, and spreads like wildfire."

Malcolm Gladwell,

The Tipping Point:

How Little Things Can Make a Big Difference

Now that we have proven the correlation between Network Gossip and Diffusion Centrality, we need to substantiate the model by testing whether or not it picks up solely the data we are interested in, or if it is either lacking or contaminated with additional information and variable effects. We can test this by running regressions. Banerjee et al. (2016) did just this, and therefore, I will use their model and data a reference when describing how regression enhance our understating of the Diffusion Centrality and Network Gossip models and their relationship.

Our Network Gossip model does not fully recognize the reasons for why our survey participates nominate other individuals in the network. Instead, we assume that an individual was solely nominated as a result of our survey participants' ability to identify an individual in the network through the remembrance of particular information (”gossip”) rather than any other characteristic such as leadership status, geographic position, etc. By running regressions, we can test whether the data calculated in the Network Gossip model is unbiased, meaning the data and values are truly depictive of one's diffusion centrality regardless of other qualities such as their leadership status within the network and their geographical location within the network etc. Should the model have an omitted-variable bias, we might conclude that we need to refine the data collection method in the model to account for the omitted variables or there exists a more efficient and accurate model by which we can collect diffusion centrality data.

To test and verify these results and this assumption, Banerjee et al (2016) use the data collected in their studies based in Karnataka, India to run regression analysis to test for potentially omitted variables. The other variables used in the regression model include demographics such as leadership status, geographic position, and caste controls, and village fixed effects.

7.2 Understanding the Model

To run the regressions, Benerjee et al. (2016) use a utility model. $u_i(j)$ is equivalent to the value/utility individual i receives from nominating individual j as a network influencer for gossip. The probability that individual i chooses individual j is such that individual i maximizes his/her utility. In particular, we express the utility model $u_i(j)$ as:

$$u_i(j) = \alpha + \beta^{'} x_j + \gamma^{'} z_j + \mu_v + \epsilon_{ijv}$$

Before breaking down the model, let us explore this concept of utility a bit more as there is an important reason that Banerjee et al. (2016) chose this until for their regression instead of another dependent variable. The utility being observed here is that which is gained through communication. Remember a participant can only nominate another individual if they have communicated with at least one other person in the network and they recall that the nominated individual was the original source of this news. The assumption here is that if an individual j has a greater influence in the network and is a central source of information within a network, individual i would find it more beneficial to talk to individual j than to another less influential individual k. Accordingly, individual i receives a greater utility from speaking with individual i over individual k. This model, however, does not directly capture the utility gained by individual i from hearing information sourced out by individual j. Instead, the model captures individual i's utility gained from mentioning individual j as the source of information. In other words, we look to find by how much each variable in our regression model accounts for the reason that an individual i choose to nominate another individual j under the assumption that individual i receives a greater amount of utility from his/her relationship with individual j than with another individual within the network.

With this in mind, let's expound the model described above. α is a constant value representing the y-intercept and β and γ are the coefficients for their respective variables. x_j is a vector of network centralities for j for the three different expressions of centrality that mentioned previous, Degree, Eigenvector, and Diffusion Centrality. z_j is a vector of demographics for j, including their leadership status, geographic position, and caste control. μ_v is a village fixed effect determined for each village by Banerjee et al (2016)'s operationalization of geographic centrality.

This fixed effect looks at multiple geographic data within each village, finds the difference between the data for each village, and normalize the difference so that they are negligible. Essentially, the fixed effect "cleans" the data, removing consequential omitted variable bias that result from unique village properties, and allows us to look strictly at the effect of centrality and demographic variables on utility. To calculate the fixed effect, Banerjee et al. (2016) first operationalized geographic centrality the inverse of each individual's distance from the village's center of mass. The accomplished this through the use of a matrix, comparing each individual's geographic distance relative to the others in the network, $\frac{1}{d(\cdot,\cdot)}$.

Lastly, $\epsilon_(ijv)$ is the associated error term for the regression model and is a Type-I extreme value distributed disturbance. Properties of this error term are conditional to large sample data and the existence of a maximum number of potentially nominated individuals, n-1. Additionally, the error term must be independent of individuals' utility and the choice behavior that affects their utility. In other words, cannot affect $u_i(j)$, $E(u_i(j), \epsilon_(ijv)) = 0$ and $E(x_j, \epsilon_{ijv}) = 0$[1]. Our regressions will test whether these properties are violated.

If our regressions show a positive covariance between any two variables, we know that the coefficient values for those variables are not picking up the variable's true effect on the dependent variables. In other words, the independent variables are not truly independent and therefore obscure the true effect of the variable of interest. Instead, the coefficients are biased and incorporate some of the effects of another variable on $u_i(j)$.

With this regression, our goal is to find the effect of our centrality measures on $u_i(j)$ and to test if diffusion centrality characteristics motivate the nomination of an

[1]Due to the distribution of data found in the log-linear regressions of the utility model, Banerjee et al. (2016) use a Type-1 extreme value distribution disturbance. This error model comes from Gumbel Distrubtuion and will find the minimum error associated with the associated log-linear regressions. The Type-I extreme values distribution disturbance model uses a probability density function to track the probability of extreme events occurring. Essentially, this error term captures residuals that result from probability of extreme events occurring. Typically, such a model is applied to engineering, finance, and earth sciences [42]. For example, one would use a Gumbel Distribution to test the probability that an atypically large hurricane hits the Eastern Coast of the US or that a once in a thousand years snow storm hits South America. In our case, the error looks at the probability that with a large sample data network saturation happens unusually quickly.

individual independent of his/her other demographic qualities. We expect to find that $Cov(x_i, z_i) = 0$ if our diffusion centrality measurements are truly independent and unbiased. If in fact, $Cov(x_i, z_i) = 0$, we conclude that our Network Gossip model estimates Diffusion Centrality and that survey participants nominate others within the network solely based on that individual's diffusive capabilities.

There are two simple ways in which we can test the covariance between variables. First, we can simply regress the variables against each other. The coefficients will tell us whether there is a positive, negative, or nonexistent relationship between the two regressed variables. Second, we can regress our dependent variable, utility, on various combinations of the independent variables used in the model. Banerjee et al. (2016) use the later method. In doing so, they first regress utility on each variable individually. Next, they run multiple regressions of utility on Diffusion Centrality paired with one of the other centrality and demographic variables. They then assess the change in the Diffusion Centrality coefficient to determine if the value from the first regression picked up an effect from the paired variables.

To do so, Banerjee et al. (2016) run both Ordinary Least Squares (OLS) and Poisson regressions for both bivariate regressions over all variables and regressions on Diffusion Centrality paired with the other variables in the model. OLS is a linear regression model and therefore, produces a linear estimated best-fit line from graphical trend in the data to predict the effect of each variable on $u_i(j)$. Coefficients for the independent variables are the estimated effect of independent variables should the dependent variable increase by one. The values are given by the best-fit line and assume a linear relationship. With OLS, the estimated model is found by minimizing the sum of the squared residuals using the software STATA for expedited calculations. Residuals are the squared differences between the vertical distance of each data point and the best-fit line. The sum of them gives us the model's error term.

The simplicity of this OLS regression makes it great for quick, comprehensible analysis of the data. However, because it is one the most basic regression method, OLS is subjected to many inaccuracies. The method assumes a linear relationship between the dependent and independent variables, which may unknowingly muddle the true understanding of the empirical data when researchers do not have a large enough data set. For example, think of the horizon it appears flat and linear, but with a large enough perspective we know that it is in fact curved. Similarly, relationships in data that may seem or appear linear may in reality vary at some unknown time period or increment. Some other such inaccuracies include, nonlinearity, biases, heteroskedacity, incorrect, non-normal error function etc. [10]. Hence, Banerjee et al. (2016) only include the OLS measurement as a point of comparison for the Poisson Regression.

Poisson Regressions on the other hand, more appropriately identify the relationships between our variables. The diffusion of information and subsequently, the reverse, nominating individuals who spread information, are derivatives of choice behavior. This means that for each time period individuals can decide between two choices: to spread or withhold information or name or forgo nominating an individual for each model respectively. These diffusion models, Diffusion Centrality and Network Gossip, capture this choice behavior for every individual in the network with respect to every other individual in the network for each time period in which the network is observed. As a result, we end up with a lot of data, but this data is not normally distributed. Rather, recall our discussion in the section titled "Eigenvector Centrality: Diffusion at $T \to \infty$" about the rate of diffusion over time.

The S-curve in Figure 7.1 shows how the rate of diffusion changes over time and as more individuals hear the information. In order for individuals to hear information, there must first be a different individual making the decision to pass along the information. If the rate of diffusion increases, we can conclude that a greater number of individuals in the time period chose to pass along the information. As such, we create a corresponding graph, Figure 7.2, that shows the number of individuals who face the decision to share or withhold information for each time period.

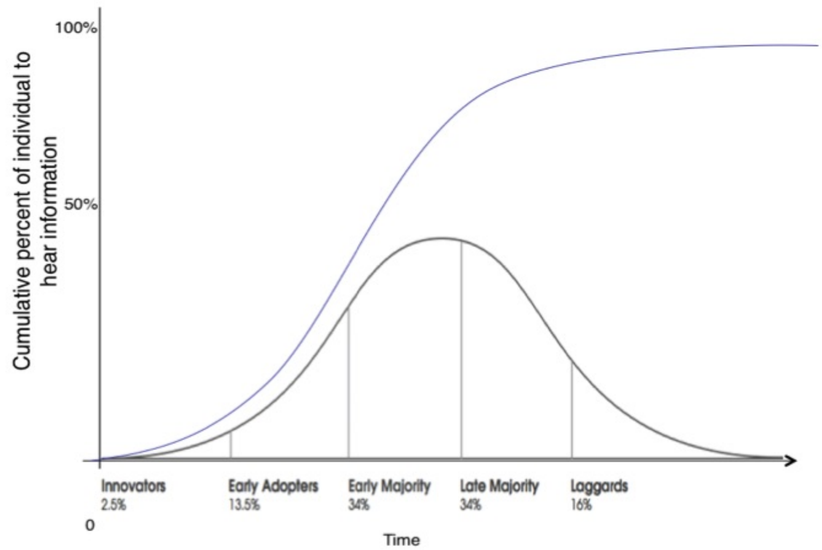


Figure 7.1: Diffusion S curve graph; The Y-Axis is the cumulative percentage of nodes who have heard the information, and the X-Axis is time over which the information spreads and stays relevant within the network.

Banerjee et al. (2016) found that characteristic of the rate of diffusion and network saturation the curve depicting the distribution of individuals making a decision was skewed rightthe data follows a Poisson distribution. The Poisson random graph model describes the probability that a number of binomial events happen over a fixed interval of time occurring independently of time [25]. Thus, graphically, data is distributed over the number of times an event is expected to occur. For Banerjee et al. (2016) and in my study, an event is the requirement to make a decision to spread/not spread or nominate/not nominate. During different periods, it is expected that a varying number of individuals within the network will make this decision and thus participate in the event.

The Poisson Regression looks at a log-linear[2]relationship between the dependent and independent variables, transforming the data (orange curve) so that it is normally

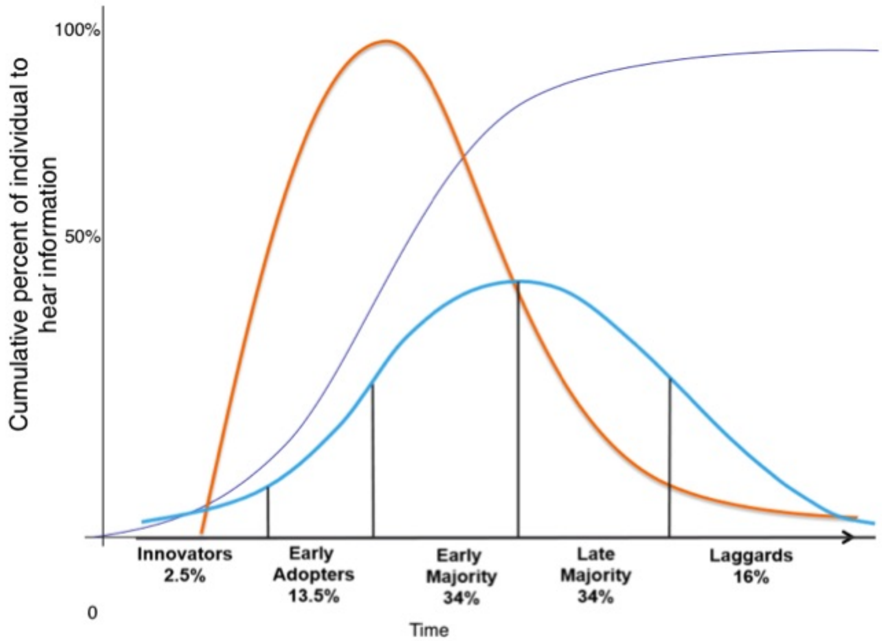


Figure 7.2: This is the same graph as that represented above with the addition of an example Poisson Distribution curve; Poisson distribution curve (orange), Diffusion S curve (Navy), Adaption curve (blue).

distributed and appears similar to the Diffusion S curve (navy) in Figure 7.2. The coefficients obtained from this regression tell by how much the log of the dependent variable changes when we increase the independent variables by one percent [23]. Specifically, for Banerjee et al. (2016)'s utility model the values generated by running this Poisson Regression are the expected number of times that individual i nominates individual j to maximize his utility.

Both the OLS and Poisson regressions articulate that of the all the centrality models considered, Diffusion Centrality is the most closely related to why individual i nominates individual j (See Appendix E). We understand this through interpreting the regression coefficient for Diffusion Centrality relative to the regression coefficients found for the other centrality methods. The bivariate regressions for both OLS and

[2] Log-linear regressions transform right-skewed data So that it is normally distributed about the average. The coefficients calculated in the regressions are consequently a change in the log of the dependent variable for every 1% increase in the independent variable. This transformation allows us to articulate dependent and independent relationships rectilinearly over time.

Poisson give Diffusion Centrality a coefficient of .285 and .607 at a 1% significance level respectively the highest, and suggestively, the most impactful of all other centralities nested in the x-variable. In other words, for every nomination that individual i makes Diffusion Centrality accounts for 60.7% of the reason for that nomination.

Although Diffusion Centrality is greater than the other centrality measures in the bivariate regressions, Eigenvector Centrality shows .605 relations to nomination utility. These results make sense since we showed previously how Diffusion Centrality and Eigenvector Centrality overlap at $T \to \infty$. Similarly, in the multivariable regressions for both OLS and Poisson, we see that again, Diffusion Centrality captures the reason for which individual i nominates individual j by the greatest percent. And again, we notice how Diffusion Centrality and Eigenvector Centrality are related measures since, when Banerjee et al. (2016) regress utility on Diffusion Centrality and Eigenvector Centrality, the Diffusion Centrality Poisson coefficient drops from .607 to .354 with Eigenvector accounting for .283 nomination utility[3].

Through these regressions we also find that although Diffusion Centrality accounts for a large percentage of individual i's reason for nominating individual j, the leadership status of individual j also greatly influences individual i's decision to nominate. The leadership status of individuals involved in the network carries an important concept even beyond that of its significance within Banerjee et al. (2016)'s regression analysis. There are two reasons that individuals nominate a leader: 1. That individual truly interacts with the leader as a prominent source of information, or 2. The individual simply expects, because of authority biases, that the leader should

[3]To further confirm the relationship between Diffusion Centrality and Eigenvector Centrality, I download Barerjee et al. (2013)s available data. They had created variables $Diffusion_Centrality_Leaders$ and $Eigenvector_Centrality_Leaders$. These centralities are only for the individuals in their study who are categorized as leaders. This, however, does not affect the correlation between the two centralities, since both variables are created from the same sample. I regressed $Diffusion_Centrality_Leaders$ on $Eigenvector_Centrality_Leaders$ to find a 58.94 coefficient and relationship between the two measures. In other words, Eigenvector Centrality determines 58.94% of an individuals diffusive ability at a 95% confidence interval. Again, since Diffusion Centrality and Eigenvector Centrality correlate only at $T \to \infty$, we expect that the relationship between the two measures would not be a perfect, 100% parallel.

be a prominent source of information. Through the bivariate regressions, Banerjee et al. (2016) shows that the leader variable's coefficient is .422 at a 5% significance linear and .868 at a 1% significance level for OLS and Poisson respectively. When Banerjee et al. (2016) run the Poisson regression with both Diffusion Centrality and Leader, we find that Diffusion Centrality only captures .553 of the nomination utility at a 1% significance level, while at a leader describes .541 of the nomination utility at a 5% significance level. The differences between the 1% significance level for Diffusion Centrality and the 5% significance level for Leader, suggests that we are 99% confident that a individual j's diffusion centrality impacts an individual i's choice to nominate and only 95% confident that an individual j's leadership title impacts an individual i's choice to nominate[4]. Accordingly, we can conclude that individuals' Diffusion Centrality and ability to influence information spread across a network best represents the reason for why other individuals nominate them in the survey and research study, but also that an individual's leadership status may, in part, determine whether or not they are nominated in our second survey.

[4]With the data provided by Banerjee et al. (2016), I was unable to regress $Diffusion_Centrality_Leader$ on Leader since the first is a derivative of the second. If I were to do so, I would find a significant relationship between the two rather than effect that ones leadership has on their Diffusion Centrality (i.e. the effect of the variable Leader on Diffusion Centrality for the general population).

Chapter 8

Methodology

8.1 Study Design and Selection of Participants

My study has been designed to accomplish two things. One, I seek to evaluate the adaptability of Banerjee et al. (2016)'s Network Gossip model to my own research around LMU's environmental sustainability communication network. And two, to identify actual flow of information and communication network around LMU's ES in hopes that it might provide deeper understanding of network strengths and weakness when trying to advance the university's ES leadership.

I surveyed 41 individuals for data from which I built a formal and an informal network. I then applied different centrality measures in hopes of uncovering individuals crucial to the spread of LMU's environmental sustainability information at the university. As mentioned in "Regressions," Diffusion Centrality does not explain fully the spread of information about a network from nominated nodes (or individuals). Rather, Banerjee et al. (2016) shows that leadership status impacts survey nominations. To account for this, I compare the informal network constructed from the survey response to that of LMU's formal organization network structure constructed from individuals' titles and hierarchical placement.

Primarily, I reference *Gossip: Identifying Central Individuals in a Social Network* to model and build my own research. The design that Banerjee et al. (2016) used to collect the data comes from their earlier work, *The Diffusion of Microfinance* (2013). Accordingly, my study design is an adaption of their 2013 work while the analysis comes largely from their 2016 article. This is the method that I adapt loosely to accommodate for the limited number of relevant and suitable participants at LMU:

The sample used in Banerjee et al. (2016)'s analysis in *Gossip: Identifying Central Individuals in a Social Network* is the same sample that was used during their initial approach to network analysis in *The Diffusion of Microfinance*. They approached 75 different distinct villages in India and identified leaders within the communities. These leaders consisted of individuals with roles and titles such as priest, teachers, bankers, and councilmen. From these titles, Banerjee et al. (2013) were able to conceive a formal hierarchy within the community. Once they identified these individuals, they reached out to households, which they picked randomly within each village. Within these households, they spoke to the head female of the household, their husband, and any other woman over 18 years old and their respective husband. These surveyed individuals then nominated (i.e. identified) other individuals with whom Banerjee et al. (2013) followed up to repeat the survey process. From these nominations, Banerjee et al. (2016) built their Network Gossip network and model. The researchers ran calculation on data collected from each village separately and then amalgamated the data to test for overall effects such as whether or not leader status effects one's Network Gossip centrality value over their true diffusive capabilities.

For my own study, the core building blocks of the network are self-evident: the university itself, which defined the outer parameters of the network, the administration, the various colleges and staff departments. My next integral task was to select individuals who do or could influence in any manner the communication and decision making of the ES concept within the network. When I applied to Loyola Marymount University's Institutional Review Board for approval to conduct research with human participants, I was asked to reach out to Rebecca Chandler about engaging campus faculty and staff in non-anonymous surveys. I was able to secure a meeting with Ms. Chandler. In addition to approving my research, she offered to provide me with resources from which I could build my initial potential participant list and an additional list of suggested individuals to whom I should reach out for the first part of my survey.

Ms. Chandler provided me with the following links: `http://studentaffairs.lmu.edu/housing/studenthousing/aboutus/sustainability/#Policy`,`http://www.greenreportcard.org/report-card-2011/schools/loyola-marymount-university.html`, and lastly a link that she noted has relevant but a bit outdated information, `http://tinyurl.com/juy2fwg`[1]. April Sandifer also suggested I look at LMU's organization hierarchical chart:

Additionally, through my roles at CURes and the Sustainability office and with Dr. Herreiner's assistance, I was able to further identify prospective ES-involved individuals. With the help of these resources, I began to identify staff and faculty who based on their formal position title and job description at LMU were likely "leaders" involved in ES initiatives. If I found a title that suggested some sort of ES engagement, I followed up by searching for other individuals with whom they work or oversee.

I established an initial list of potential participants, faculty, staff, and students at LMU who, from their title and publicized work, I thought were potentially involved in LMU's environmental sustainability. To that initial group of people, I sent part one of my two-part survey model. This first survey served as a baseline survey, consisting of 15 questions from which I could gather information on each individual's self-perceived

[1] Do to website reconstruction, this url is no longer active

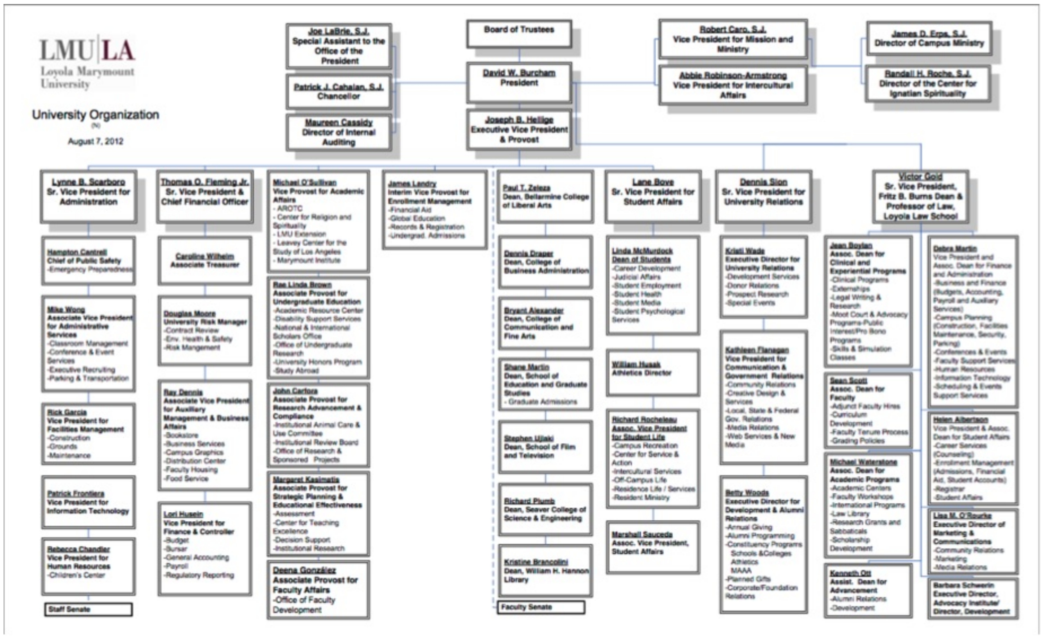


Figure 8.1: LMU's 2012 University Organization Chart.`http://intranet.lmu.edu/Assets/Administration+Division/Human+Resources/University+Org+Chart.pdf`.

involvement in LMU's ES. (See Appendix F). I administered this survey via PDFs sent by email and a hard copy mailed to each individual's internal LMU mailbox. Those who responded either did so by emailing me a completed electronic copy or mailing a completed hardcopy to my own LMU mailbox. I heard from 25 of 41 of individuals to whom I sent the first survey. Of the 25 only 3 did not qualify for the second round of surveys due to responses exhibiting little to no involvement in any area of LMU's ES efforts. Consequently, my sample size for second survey and from which to collect crucial data dropped from 25 to 22.

Due to this small sample size, my study differs in significant respect from Banerjee et al. (2013). If respondents mentioned other individuals, I recorded that reference. If an individual mentioned was not one whom I had already sent out the first survey, I immediately emailed the first survey to the new individual. The main distinction between this method and that of Banerjee et al. (2013) is that they used their first, baseline survey to generate characteristic data such as village leadership, geographical features, the existence of NGOs, and self-help groups about the different participating villages. My study, however, considered only one community, LMU, and information about LMU analogous to that identified through Banerjee et al. (2013)'s first survey is available publically through LMU's website. Instead, I constructed my first, baseline survey to gather characteristic data about individual's prescribed role at LMU and within their particular departments. This helped me to also understand better and validate the departmental hierarchy information that I gathered from LMU's website, `www.LMU.edu`.

When entering the data and open responses from the first survey, I noted an additional three individuals who were mentioned but who were not a part of my initial list. I then sent these individuals the same survey. Some participants mentioned names of individuals with whom LMU partners. However, I excluded these individuals from the study so as to maintain a more closed network, and concluded with a sample of 41 recruited individuals. Of the 41 individuals, 25 responded, and only 22 of the 25 provided responses that suggested a moderate to strong role in the discussion and decisions of ES at LMU whether in education, maintenance, planning, implementation, or management.

Accordingly, I reached out to the 22 qualified participants, requesting their participant in the second survey (See Appendix G). Initially, I had planned to conduct the second survey as an in-person interview in the same manner as Banerjee et al. (2013). However, due to limited time at the end of my senior semester and the complexity of trying to schedule interview times with participants, I resorted to sending it once again by email. This time, however, I built the survey on Qualtrics, limiting participants to electronic participation since I did not also mail out a hard copy.

The second survey consisted of 24 questions designed to collect social network data. I adapted these questions from Banerjee et al. (2013)'s survey module and designed them to capture different dimensions of individuals' social behavior and psychology from which I could build and test LMU ES's informal network structure. These questioned probed: names of individual from whom ES information is sought, the extent of involvement of these individuals in LMU's ES, areas of ES in which these individuals are involved, areas of LMU ES in which they themselves seek information, names of those who implement ES practices at LMU, names of those who advertise information about LMU ES, names of perceived ES leaders at LMU, and names of LMU departments most involved in various aspects of LMU's ES efforts.

Again, if a participant mentioned the name of someone not yet a part of the study, I sent him/her an email with the first survey and requested his/her participation. This happened only once. Upon receiving their responses, I sent them the second survey. The method for recording responses from the second survey comes from Banerjee et al. (2016) since Banerjee et al. (2013) only used the data insofar as it provided parameters and demographic controls for their regressions. Accordingly, individuals that respondents mentioned in their survey responses were recorded as nominations for those who might be diffusively central. Departments that respondents mentioned were recorded as well so that in the future I might be able to run centrality measurements across the departments as Banerjee et al. (2016) did with the households. In the end, I only received 15 responses from the 23 recruited individuals thus, leaving me with a sample size of 15 from which I could test and calculate the informal network centrality measurements.

Once I collected both surveys' data, I constructed both graphically and matrix representations of LMU's formal and informal networks. I imputed the matrix into MatLab to run the data through two different versions of Diffusion Centrality code. The first code, I wrote with the help of my friend Bryce Currey, and the second code, I sourced from MIT's "Overview of metrics and their correlation patterns for multiple-metric topology analysis on heterogeneous graph ensembles." This first code makes use of Diffusion Centrality's correlation to Eigenvector Centrality at $T \rightarrow$

∞. I used the network graphs to enhance my quantitative results and to address network behavior and characteristics that my survey and quantitative data may have overlooked or omitted. To access the recorded data files, use this link:

8.2 Limitations

The small sample size lead me to believe that the results for this second survey may not be as conclusive as I would have otherwise preferred. I addressed this issue with Dr. Herreiner who suggested that this sample size, particularly for how limiting our study parameters are, would not be problematic. Rather, I found that my biggest set back came not from the small sample size, but instead from my failure to adapt important nuances in the design second survey questionnaire, leading to notable data limitations. I believe gaps in my survey questions affected my Network Gossip data from the second survey, and thus, caused the Network Gossip values to inaccurately depict individuals' relationship insofar as they relate to the communication environmental sustainability at LMU. The results from my second survey, although they may not capture the intended information, are interesting nonetheless. So, I will still discuss these results and their implications "Findings and Discussions of Research Outcomes" along with the survey and centrality results for both the formal and informal networks.

Again, the concepts and methodologies I use from Gossip: Identifying Central Individuals in a Social Network to build the second survey's questions are derived from the empirical and simulated data analyses in *The Diffusion of Microfinance*. My approach to the data and analysis, however, come from Gossip: Identifying Central Individuals in a Social Network because of its approach to building an informal communication network from collected data without having to first seed and track a piece of information.

As I mentioned in the previous section, the most critical element to understand and adapt from Banerjee et al. (2016)'s study is the survey questions. Without thoughtful composition of the survey questions, survey participants may not provided the appropriate data from which to identify diffusively central individuals. Instead, without the correct survey questions, I risk breaking the correlation between Banerjee et al. (2016)'s Diffusion Centrality and Network Gossip models. Thus, I learned that the most critical element for a successful adaptation and replication of this study is not the calculations, but instead gathering and generating the appropriate data.

Although I administered the survey questions to very different groups of people and under a very different context than Banerjee et al. (2016), I needed to frame and word such the vernacular generated a similar response and conveys the same objective. I constructed the questions for the first survey with the objective of gathering more information about each individual's formal role at LMU and with LMU's ES initiatives. The questions helped me to build a network representing formal, hierarchical and job description driven relationships among faculty, students, and staff participating in LMU's ES as well as learn who might be candidates for my second survey. I developed the second survey's question with the intention of following the structure and sentiments of Banerjee et al. (2016)'s 12 survey questions. In doing so, I hoped to collect data about the informal social relationships within the network irrespective of formal organizational ties.

Since the formal structure of the university is public knowledge, the first survey simply served as a reaffirmation of information available through LMU's website and HR resources. Thus, first survey successfully captured the appropriate information. The second survey proved more complicated and consequently captured the appropriate data with much less success. Banerjee et al. (2016) asked 12 questions: names of those who visits the respondent's home, those whose homes the respondent visits, kin in the village, nonrelatives with whom the respondent socializes, those from whom the respondent receives medical advice, those from whom the respondent would lend money, those from whom the respondent would borrow material goods (kerosene, rice, etc.), those to whom the respondent would lend material goods, those from whom the respondent gets advice, those to whom the respondent gives advice, and those with whom the respondent goes to pray (at a temple, church, or mosque). These questions got at the breadth of the respondents involvement within the village community and the dimensions about which the respondent socializes with others within the network. I modeled my questions accordingly, asking: (i) names of three individuals to whom the respondent would approach for information regarding LMU's environmental sustainability, (ii) to rank on a 5-point scale the extent to which each nominated individual is involved in LMU's environmental sustainability efforts and in which areas of LMU's environmental sustainability each individual is involved, (iii) for which areas of LMU's environmental sustainability the respondent seeks the most information, (iv) three individuals the respondent recommends working with for the implementation of environmental sustainability practices at LMU, (v) three individuals the respondent feels are influential in disseminating information around the university about LMU's environmental sustainability efforts, (vi) three individuals the respondent considers to be leaders in the LMU's environmental sustainability efforts, (vii) the department the respondent feels plays a fundamental role in advancing environmental sustainability practices on campus, (viii) the department the respondent feels contributes the most to the education about LMU's environmental sustainability, (ix) and the department the respondent feels best oversees and preserves environmental sustainability practices on campus.

The first difference between the survey questions from Banerjee et al. (2016) and my study is that I ask the respondent to not only think of particular individuals, but also different departments involved in ES initiatives. I did this to observe if network central individuals were correlated with network central departments. In other words, I wanted to learn whether communication around LMU's ES happened through persons' individual efforts or through departmental efforts as a whole. T he second difference, and perhaps the most detrimental to my data, is that my questions are phrased in such a way that they lead respondents answer based on sentiment and opinion, who they "feel," or "believe." Phrasing such as "which is" or "who is" the "best" prompts the respondents to give an opinion, meaning their responses might be based on theoretical communications rather than actual practice and interaction with the nominated individuals or departments. For example, when asked, "which LMU department plays a fundamental role in advancing environmental sustainability practices on campus," one respondent answered, "The Sustainability Office-but not at the level that is suggested. They are supposed to coordinate all of the programming that is developed by ALL departments on campus who are interested in sustainability." This respondent concluded by selecting only "slightly influential" when asked how effective the department is in caring out their role. These answers, elicited by the structure of the question, are opinion-based rather than hard, testable fact.

Third and last, I recorded all name mentions from both the first survey and second survey as undirect links, meaning that even if only one respondent mentioned a particular individual and that individual did not also mention the first respondent in their survey, I recorded the data as though the relationship is reciprocal. I recorded the data in the manner under the assumption that respondents nominated individuals with whom they interact. Therefore, even if respondent a nominates another individual b, but individual b does not nominate our first respondent, I assumed that there does still exist a relationship between the two people. This is a safe assumption for my first survey. If a respondent of the first survey names another individual as their boss, then there does undoubtedly exist and undirect, reciprocal relationship between the boss and employee. This also allows me to say with confidence that the formal network is symmetrical.

For the second survey, however, I may have assumed falsely an undirect relationship between the respondent and their nominations. This error is a result of the improper phrasing and structuring of the second survey's questions. The questions are constructed words such as "would" and "think," leading me to believe that the respondents most likely answered based on opinion. Therefore, I cannot say with confidence that the relationships are undirect. The respondent might have nominated an individual who they learned is involved in LMU's ES efforts, but from whom they have never received any communication or information. For example, when asked if there is a particular department that oversees ES initiatives, one survey respondent exclaim, "it should be the Sustainability Office, but it is not."

Furthermore, I know that many of the individuals in the informal network have leadership roles within their department. There formal job title and job description grant them the ability to make and approve decisions. Consequently, there exist the possibility that my data is only picking up network expectations based of job title and leadership status (coherent with the formal network) rather than actual informal network communication. Additionally, emailing my surveys and using primarily an online survey method for my research may have thwarted the quality of my data. Nonetheless, the responses and lack thereof that I received using this method provide further grounds for discussing how to best communication and progress environmental sustainability information around LMU. For the first survey, I gave participants the option of responding either via email online or through the mail using a hard copy of the questionnaire. For the second survey, I restricted responses to Qualtrics, an online survey platform accessed through email. I was feared that restricting the method for participation in the second survey would reduce the number of response I received. However, I got a better response rather with my second survey. Of 41 individuals received the first survey and 25 completed the questionnaire, giving me a

61% response rate. 23 individuals received the second survey of which 15 responded, giving me a 65% response rate.

Instead, the problem with my procedure is not a matter of email vs. mail, but instead an absence of in-person discussion. I spoke with an individual who informed me that they completed a hard copy of the first survey and that they survey, completed and ready to be sent, sat on their desk. However, I never received their completed survey even after a few requests. I spoke with another individual, requesting their completion of the second survey several times over the period that the survey was open. At each request, they responded that they fully intended to participate and that the survey was open in their email. Still, I never got a response from this individual. Consequently, I did not collect data from these individuals and believe that their response might have been influential in my research, providing a more thorough set of data from which to run my centrality measurements. While these were only two particular individual to whom I spoke, I conceive that other potential participants who might have given meaningful data, forwent the task of returning a completed survey. I theorize that this is perhaps similar to Cohen et al. (1972) proposition that employees of the university are stricken with an increase in load of work that leave less attention-grabbing activities overlooked and forgotten. Consequently, comparing the study results between those found to be diffusively central (whether or not the findings result from actual node characteristics or simply from study participants' expectations) with those who participated and respond add another interesting topic for discussing the communication and successful action for ES initiatives at LMU.

Chapter 9

Findings and Discussion

9.1 Graphical an Quantitative Results and Observations: Formal Network

Using the information collected from the first survey and other online LMU resources, I built a formal network of faculty, staff, and students whose job title and role at LMU includes them in campus environmental sustainability initiatives and decision. As such, the network depicts relationships based off the expectation of involvement and attentiveness carried both by individuals within the network themselves and by individuals outside the environmental sustainability domain and even LMU.

The colors of the nodes correspond to an individuals participation within the study, and the colors of the lines refer to how I established the connection.

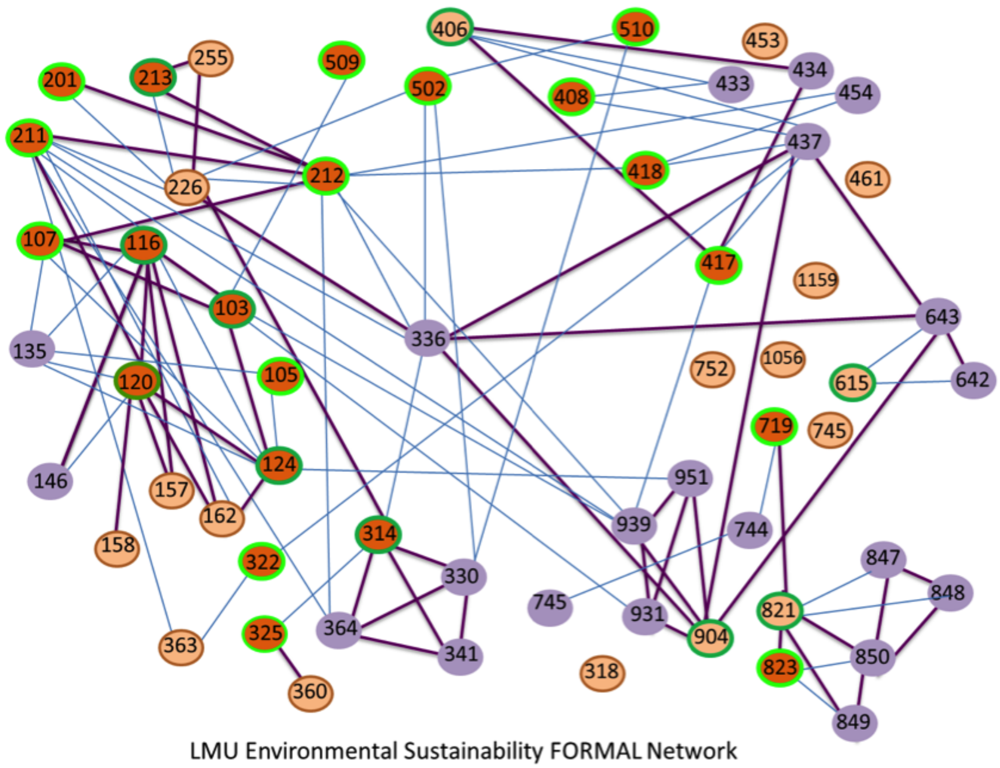


Figure 9.1: LMU formal network graph

Color	**Representation**
Light Orange	Individual was asked to participate in the studys first survey.
Dark Green Outline	Individual completed the studys first survey.
Dark Orange	Individual was asked to participate in the studys second survey (having first completed the first survey).
Bright Green Outline	Individual completed the studys second survey.
Blue Link	Connection was established through the responses in the

The hundreds number corresponds to specific LMU departments, and the number following is simply an individuals identification number.

Department Name	**Department Code**
Facilities Management	1
Center for Urban Resilience (CURes)	2
Seaver College of Science and Engineering	3
Bellarmine College of Liberal Arts	4
External Organization Campus Representatives	5
Hilton School of Business Administration	6
Housing	7
ASLMU	8
Administration	9
College of Communication and Fine Arts	10
School of Film and Television	11

Table 9.2: Formal Network Graph Department Code

I transposed this network shown in Figure 9.1 into an adjacency matrix, and then used MatLab to run Diffusion Centrality calculations as it corresponds to Degree Centrality and Eigenvector Centrality on our formal network. Calculating Degree Centrality, I find that individuals 212 with (11) links, 116 (9), 211 (9), 226 (8), 120 (8), 124 (8), 336 (8), and 437 (8) are the 10 most Degree Central individuals within the network.

Next, I tested for Eigenvector Centrality, using two different coding methods. First, Bryce Curry helped me to code Banerjee et al. (2016)s benchmark parameters to find Eigenvector Centrality. In this method, E[Diam] = 8 and $E[\frac{1}{\lambda_1}] = 1/6.0515$. Second, I used a simple code specifically for Eigenvector Centrality written for MatLab by MITs Strategic Engineering group [6]. (See Appendix H & I for codes). This second version confirmed that the estimated diameter, 8, was an appropriate benchmark time period for $T \to \infty$. The $E[\frac{1}{\lambda_1}]$, λ_1 being the largest right-hand eigenvalue, was the same in both models. The two methods showed similar centrality distributions about the individuals in the network; the main difference between the two methods is that the second presents individuals centrality as a percentage value with respect to the other individual in the network whereas the first simply states by how much an individual is diffusively central. Using the second (first) method, I found that nodes 212 and 211 are the most central with values .3398 (13.9976) and .3285 (13.8090) respectively. The next most central node is 116 with a value .3141 (12.0680.) Unsurprisingly so, the least central nodes are those that are disconnected and isolated from the networks giant component. Generally, I found that most of the nodes with the lowest centrality values are from BCLA, department number 4. Conversely, the data suggests that department 2, CURes, hosts the two most diffusively central individuals within LMUs formal ES network.

I must admit that I was quite surprised to learn that the most central figures in this network come from CURes. Two explanations might account for this finding. First, although the majority of CURes work applies to external research, that is research about environments and with organizations external to LMU, they secure a lot of their resources to carry out the research from LMU particularly, LMUs human and social capital. Every study conducted and every effort made by CURes is successful in part because of the time, knowledge, and physical support contributed by LMU students, faculty, and staff. CURes hires student interns to help with research like the assessment of the social value of the Baldwin Hill recreation center and the biogeographical research done at the Ballona Wetlands and faculty members help to analyze the collected data before redistributing the resulting to partnering organizations. In order to enlist the help of persons outside of CURes and as a part of their job, CURes employees must create out reach and build connections with departments all across campus. Furthermore, many of the CURes staff involve themselves within the LMU community by guest lecturing and teaching various classes. Thus, their role as lecturer for whichever department increases the degree of their formal network connection.

Second, since I, myself, work in CURes, I may have unintentionally given individuals within CURes more connections due to easier and deeper research about to whom they report and with whom they work as required by their job description. While this is likely, I must argue that not every individual in CURes is valued above the others in the network, and so, this might just be coincidental. Perhaps, however, theses CURes individual might have also provided more detail in their first survey responses, giving me a greater depth of understanding of how their job connects them to the network because, in knowing me personally, they are more sympathetic to my research.

Also, while individuals in BCLA show the lowest centrality values, it is still noteworthy that there are quite a few BCLA faculty members included within the network. This means that even with low centrality, their title and others expectation of their role at LMU involves them in some way with LMU ES initiatives and education. Whereas, there are many other departments that are not highly represented, if at all. For instance, the network contains very few individuals from the business school, only one from the School of Film and Television and one from the College of Communication and Fine Arts.

By looking at the formal network graph, we can also see that on a departmental level, Facilities Management, ASLMU, and the business school are the densest. That is, they have the greatest number links to the total number of possible links (within the department). Since these connections are dictated by formal job roles and titles, the density of these departments suggests that these departments are characterized by specialization. In other words, each individual must communicate and work with all other persons in the department in order to accomplish ES tasks.

9.2 Graphical and Quantitative Results

Using the data collected from the second survey, I built an informal network. I write informal in quotes because the factors discussed in the previous Limitations, lead me to believe that this network, constructed from the undirected relationships presented in the second survey results, does not accurately depict the full scope of the informal social communication network about LMUs ES initiatives.

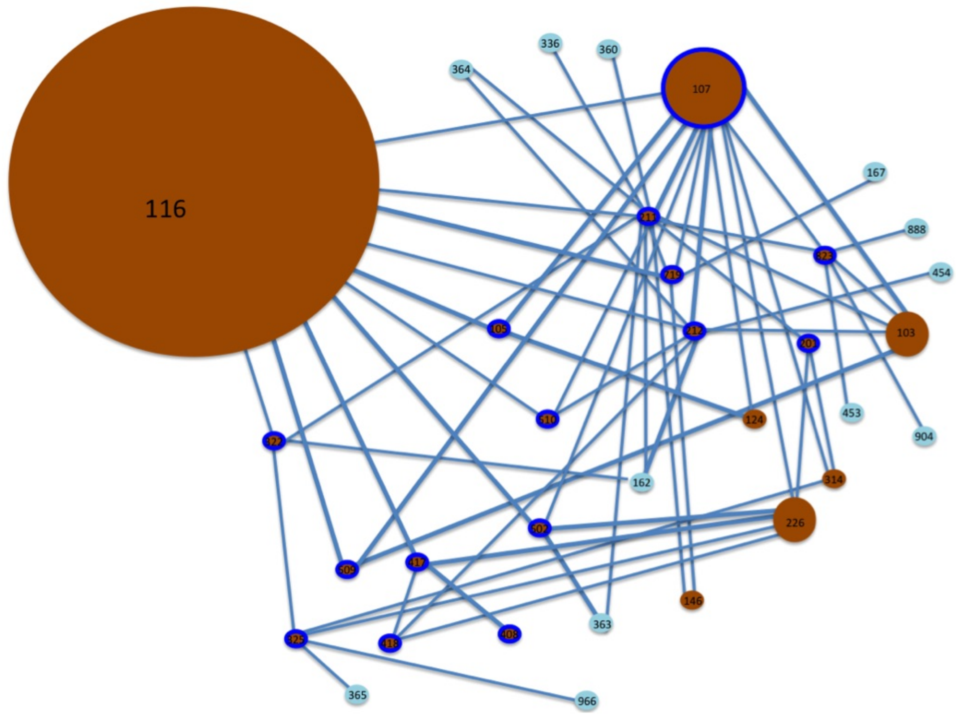


Figure 9.2: LMU informal network graph

Like with the formal network graph, the colors of the nodes in the informal network graph correspond to an individuals participation within the study, and the density of the lines refer to the type of connection. The node number codes remain the same.

Color	Representation
Dark Orange	Individual was asked to participate in the studys second survey (having first completed the first survey).
Royal Blue Outline	Individual completed the studys second survey.
Light Blue	Individual was mentioned in the second survey, but did not participate in the study.
Thin Line	Survey Participants nominated the linked individual 2 or less times in response to the studys second survey.
Thick Line	Survey Participants nominated the linked individual 3 or more times in response to the studys second survey.

Table 9.3: Formal Network Graph Color Code

Despite my hesitation with the quality of the results from my second survey, I went ahead with piecing together the network graph and running the network centrality measurement out of curiosity for what the data might suggest. I ran the data using two different versions of the recorded data.

In the first set of data, I recorded the nominations as binary undirected relationships, building an adjacency matrix of simple 1s and 0s. In other words, even if a survey participant mentioned a non-survey participate where the later did not have the opportunity to articulate their own responses and nominations, I recorded the pairs relationship as undirected. This follows the recording method used by Banerjee et al. (2016). In Figure 9.2, these relationships are simply shown by the links between individuals. Analysis for node connection (Degree Centrality) and node communication according to this method is such that if an individual is a survey participant and shows a greater number links than other individuals in the network, that participant is more connected within the network and has a greater Degree Centrality. Similarly, if an individual is a non-survey participant and shows a greater number of links than other individuals within the network, that individual is more connected within the network and has a greater Degree Centrality. If both end nodes in a link are survey participants, we assume that communication between these survey participants is initiated by either individual. If an end node is a non-survey participant, however, I infer that the survey participant rather than the non-survey participant initiates the communication between the linked pair.

In the second set of data, I weighted each relationship by the number of times the pair of individuals mentioned each other in the second survey. If individual a mentioned individual b for two, three, or four of the survey questions in the second survey, I recorded their relationship with the corresponding number, 2, 3, or 4 rather than a binary 1. This shows a stronger connection between the pairs with higher relationship numbers and suggests that there exists a greater chance that the pair communicates than a pair with fewer mentions between them. In the network graph, this relationship is shown by the thickness of the link. The size of the nodes in the network graph represent the number of nominations received by each individual relative to the rest of those in the network. Individual 116, for example, received 25 of the 121 total of votes for a 21% consideration for the second survey ES responses. Individual 211, on the other hand, only had 5 of the 121 votes and therefore, only a 4% nomination. Interestingly, however, individual 116 has a degree of 10 (i.e. is linked to 10 others within the network). Individuals 211 and 107 have degrees of 12. Therefore, 116s prominence comes not from his degree centrality, but instead from the fact that those who nominated him/her in the second survey believed that he/she plays an important role in various aspects of LMUs ES (shown by line thickness). This distinction also appears in the networks diffusion centrality measurements, which show that despite 116s distinction in the informal ES network, he/she is not the most diffusively central individual.

Using the same methods as with the formal network, I ran centrality measurements for both sets of data, the binary and weighted. For each, I found the diameter of the network in Figure 9.2 to be 7. Using the Banerjee et al. (2016)s benchmark parameters and then MITs eigenvector MatLab code[1]with the binary data, I found $q = E(\frac{1}{\lambda_1}) = (1/5.3071) = .1884$. As a result, I found that individual 211 is the most diffusively central with values 15.4815 and .4545 for the binary and weighted data respectively. Individuals 107 (14.4830 and .4262) and 116 (13.4024 and .4035) were the next diffusively central individuals within the network. Using this data, we again see that individuals from CURes and Facilities Management are the most diffusively central within the network. Should I consider strictly the results of these calculations, I would recommend seeding all critical environmental sustainability information, ideas and tasks through individuals 211, 107, and 116.

[1]Bounova, G., de Weck, O.L. "Overview of metrics and their correlation patterns for multiple-metric topology analysis on heterogeneous graph ensembles", Phys. Rev. E 85, 016117 (2012).

However, in light of my assumption made while data recording, that all nominations represent undirected relationships between individuals, and the potential collection of opinion-based responses, these individuals may not necessarily be the best at diffusing information, but instead might simply be those expected to be the most connected. Thus, these individuals may have the potential to diffuse information to the greatest number of individuals within the network, but this is entirely dependent on their willingness to actual communicate and their actual collaboration with others in the network.

The weighted data my second survey suggests that those with higher values are more likely to communicate. Using both Banerjee et al. (2016)s benchmark parameters and then MITs eigenvector MatLab code with this data, I found that individual 116 is the most diffusively central with values 18.2192 and .5080 respectively. Individuals 107 (16.5085, .4624) and 103 (12.4160, .4194) were the next diffusively central. Interestingly enough, all of these individuals are based in Facilities Management. However, because of the embedded sense of opinion in my survey responses this data is better understood in this way the greater the weight (or heavier the edge) between a pair, the more important one individual in the pair believes or expects the other to be in communicating, decision-making, and executing ES tasks. Therefore, if an individual within the network is perceived to be more important, the more we can expect others within the network will go to him/her for ES information. However, the probability of these individuals communicating has not been established by this data.

Consequently, I have yet to identify mathematically informal network communications with which I can compare to the expected communications rooted in the formal hierarchical, role-based network to identify any discrepancies and potential network breakdowns. The qualitative results, however, provide a sound jumping off point from which to address differences between expected and actual network behavior and communication. Of particular interest are my own interaction with and the network behavior of individual 116. Both quantitative and graphical results convey that individual 116 participates and employs a great deal of attention to ES communication, decisions, and initiatives. However, my interactions with 116 and other qualitative data, again suggest that my data may not accurately depict the reality of individuals network behavior.

9.3 Comparing the Formal and Informal Networks

Although visually the two networks appear quite different, my quantitative data shows that the formal network and informal network are quite similar. This may suggest one of two things. One, LMU's ES formal and informal networks are in fact closely related. Individuals speak socially with those whom their job also requires formal communication and collaboration. Two, the similarities result from my imperfect adaptation of the Network Gossip survey, which confirms that individuals' expectation and opinion about others' job role and responsibility in decision-making follows that of the formal, hierarchical network.

The formal network graph, made up of 59 individuals, has a diameter of 8. The informal network, made of up 32 individuals, has a diameter of 7. Given the 27 persons different between the two networks, diameters 8 and 7 are relatively similar, and suggest that the range of the network between the two is not very different. However, this measurement does not say anything about the communication behavior within the network. Instead, this similarity confirms that throughout my research and adaption of Banerjee et al. (2013) study design, I was able to maintain a closed network and keep my study within the context of individuals at LMU seemingly involved with LMU's ES endeavors.

Instead, comparison of the formal network's and informal network's centrality data gives insight into any similarities and dissimilarities between the prescribed organizational behavior and the de facto structure and behavior around LMU's ES. From the formal network, individuals 212, 211, and 116 are the most diffusively central. 211, 107, and 116 are the most diffusively central in the informal network. 211 and 116 appear as diffusively central in both networks. There are a couple of ways to assess this information. First, as noted before, LMU's formal and informal networks could in fact be similar. The permissive structure of universities allows for a softening of hierarchical relationships. Consequently, relationship dictated by hierarchy and job roles might also experience a softening of formality. In other words, the culture around work/life relationships between boss and employee and coworkers may not be as discrete as it is in traditional organizations. As such, this data might simply suggest that individuals 211 and 116 spend time both professionally and socially with a number of other individuals within the ES network.

Due to the small number of individuals intensely involved in the LMU's ES, I do believe that, in reality, the networks have similar characteristics. However, I would be truly surprised if, at this point in time, the two network were as identical my quantitative data suggests. This belief comes from my own biases I heard about and experienced myself a lot of communication failures, leading me to believe that the true communication pattern is quite dissimilar. And, while I know it is not enough to prove this hypothesis from a hunch, I must acknowledge that by rejecting that the networks are actually incomparable because of the limitations with my data and the possibility of inappropriate data, I must also reject the former analysis on the same account. Consequently, this issue remains largely unanswered.

Still, through analysis of individuals perception provides for a primitive understanding of individuals' communication behavior regarding LMU's ES. Individuals 211 and 116 appeared in the top three most diffusively central for both the formal and informal networks. 212, however, dropped out and was replaced by individual 107. Since 107 is not one of the most central within the formal network, it is apparent that others within the network believe that 107 is an influential and important player in all aspects of LMU's ES. Even if survey participants answered the second survey strictly by opinion, the fact that individuals perceive this to be true suggests that 107 might perform above and beyond the responsibilities and collaboration requires as prescribed by his job.

Furthermore, the graphical differences allow for meaningful analysis despite data limitations. In the formal network graph, we can see that the distribution of and communication between individuals is mostly segregated by department. In the informal network, such segregation is not as clear. Instead, we see individuals linked with other individuals irrespective of the department. This observation hold regardless of whether or not the data collected from the second survey is based on perception or actual communication behavior. As Bekessy et al. (2007) suggests, this demonstrates visually the distortion between task allocation and accountability. ES decision-making tasks may be allocated to various departments, however, the individuals within those the departments are responsible for the completion of those tasks. Consequently, it appears as though only a handful of people actively participate in the LMU's ES, and subsequently, there are only a handful of people who get held accountable for the failure of a decision. The cause for this is not determined in this study. Although, through the context of this study, I can speculate this behavior results from poor communication. In other words, although individuals may be linked and communication theoretically dense (as shown in the formal network), the probability that individuals actually communicate about and collaborate on LMU's ES is low.

9.4 Qualitative Data Results and Observations

As shown in Figure 9.2, many individuals within the network expect 116 to engage actively with LMU's ES program. Additionally shown in Figure 9.2, however, 116 did not complete my second survey. I spoke with this individual often both over email and in person. A handful of those times, I reminded them about the study and the survey, and they expressed an interest in participating. So, to my knowledge, the lack of response was not from a lack of interest in the study. Perhaps, then, as Cohen et al. (1972) suggests, this individual was besieged by an increase in workload and frequent flight of attention. This sort of response proposes that ES workload and tasks might be more effectively and efficiently addressed if they were allocated to a greater number of individuals. Again, however, the error made in the second survey question development makes it unclear whether failure in communication with me impacts the work and communication with others in the network.

Qualitative data from the open-end questions gives me further reason to believe that the data does not show the whole scope of network behavior around ES at LMU. Over half of the second survey participants did not nominate three individuals for each question requiring them to do so, and a majority of these respondents did not nominate anyone. Rather, they left these questions blank or instead responded with statements such as, "I don't know," "not sure," "I am not in a good position to answer that," "none!" and "no one." I also found, interestingly, that a few individuals whom others nominated as influential in LMU'S ES efforts, did not believe themselves that they were an important part of the network. Particularly, from the second survey, individual 408 received 4 nominations more nominations than any other individual from BCLA. This indicates that other survey participants believe 408 to be involved in LMU's ES. This individual felt otherwise.

408 elected to keep all question responses blank, exclaiming, "Sorry! The survey presupposes a body of information that I cannot assess." These short quick responses and the contradiction of 408's involvement express an underlying feeling of apprehension. The survey participants feel as though there is not or that at least they are not aware of a particular individual that really manages, oversees, or coordinates ES decisions around campus. Whether there truly does not exist a lead individual or whether there simply exists a lack of awareness about such individual on campus, this study makes clear that there is a breakdown in communication because of this unawareness. Without a knowledge of who effectively contributes to what decision and what aspect of the advancement of ES programs individuals are more likely to forfeit decisions and activities because of a lack of support, accountability, and progression. Such knowledge is attained through effective communication of individuals' activities. Consequently, we can infer that LMU faculty, staff, and administration are not efficiently communicating with one another nor utilizing the available resources and that there exists a discrepancy in job expectation and actual behavior.

Open-ended responses about LMU's departmental roles with ES show that individuals are aware of resources accessibility and positively believe that LMU still has the ability to grow and create permanent ES change. Like with individual nomination focused questions, a majority of survey participants opted out of answering the survey questions that asked about LMU's departmental roles with ES. Those who did respond, however, provided more readily detailed responses as to why they hold a certain opinion. Again, there is a diversity of answers, suggesting that there does not exist a single department that manages ES decisions on campus. One individual believes that LMU's Environmental Science program promotes the preservation and oversight of LMU's ES through academic work. Another asserts that the department primarily responsible for ES decisions "should be the Sustainability Office, but [that] it is not." Others believe that LMU's Recycling most effectively carries out ES decisions because of its ongoing and relentless "partnerships with ASLMU, RHA, LMU for Others, and LMU Family of Schools." For the most part, these responses are positive and prove that each department can and should participant in the advancement of LMU's ES. There exists within each department the resources to positive add to LMU's ES growth. Accordingly, these qualitative responses illuminate a fundamental error in the behavior and expectations of the network. Department related responses suggest that LMU's departments have the resources and capabilities to beneficially impact and reaffirm LMU's ES efforts. However, responses about individuals' involvement within the network are not so positive and suggest a lack of efficient communication within the whole of the network.

The problem uncovered here is that individuals assume task allocation to departments rather than particular individuals. Departments, however, cannot be held responsible for failures and successes. Rather, the individuals with the departments must be held accountable. Instead, the individuals must work with those in other departments to establish a system for task distribution and accountability to better articulate means for university-wide collaboration and an efficient allocation of resources and individuals' attention. Additionally, the diversity of departments mentioned in the responses suggests that the establishment, success, and management of ES efforts concern numerous departments.

9.5 Environmental Sustainability Stakeholder's Committee (E2SE)

In 2007, LMUs president at the time, Robert B. Lawton, established the Environmental Stewardship and Sustainability Committee. This was one was in which faculty, staff, admin, and students from different departments and areas of focus could convene to discuss LMUs environmental sustainability efforts and impact. This meeting allowed for cross-university collaboration on the formalization and actualization of ES decisions, ideas, programs, etc. Since the founding of the E2SE, LMU achieved a lot of environmental sustainability milestones, including but not limited to, the opening of two LEED certified buildings. Still, of recent, LMUs ES innovation and growth has dwindled. At the beginning of this paper, I mention that was intrigued to see how individuals communicated about ES without the committee. It is through my qualitative data that I learned about the impact and importance of this committeeso much so that I believe the results are important to address.

Through my second survey, one participant explained that, In theory, all of E2SE should play an active role [in LMUs environmental sustainability decisions]. In practice, however, it seems like a group of illuminati run the decisions of whos in (and by extension, whos out) of these discussions/innovations/evolutions. The respondents statement indicates that involvement and appointed attention to ES decisions is exclusive. I predict, however, that this exclusivity is not intentional, but instead a result of poor communication. I believe this in part because I have learned through my work at CURes and discussions resulting from my study that the E2SE meetings no longer take place. In fact, the committee has not met since Spring 2014. The tone and the present tense of the participants statement suggests, however, that the participant describes a current sentiment and that presently, a select few dictate who is and who is not involved. Without the meeting any exclusivity experienced ensues from a select few individuals choices to work with others whom they know will respond attentively to their efforts. After attempting to speak with perhaps formally involved individuals, these select few learn from whom they receive the most benefit for their time spent in trying to communicate, and consequently, other individuals get looked over.

While this exclusivity may be unintentional and singularly felt by this particular participant, the desire for some sort of general assembly of individuals to discuss environmental sustainability is widespread. During the course of my research, I had the opportunity to explain further my interest in doing this researchthat I chose to focus my research in the context of LMUs environmental sustainability because of my interest in the matter and my involvement with GreenLMU and CURes. During these different conversations, as addendums to their survey responses, and in the surveys open response sections, nine different participants asked, requested, and suggested that the monthly E2SE meetings resume. It is thus evident to me, that there are individuals at LMU who have an undeniable appreciation of and a demand for environmental sustainability collaboration and teamwork. And, it is clear that individuals at LMU do not undervalue environmental sustainability, but instead, barriers caused by ineffective communication due to a lack of university knowledge about specific ES initiatives and decision-making authority prevents LMU and the persons involved from continued environmental sustainability leadership. I thus recommended that LMU reinstates some sort of assembly in order to advance their environmental sustainability programs and once again become a leader in the field.

Chapter 10

Conclusion

Doing this research gave me the opportunity to examine the interactions between LMU faculty, staff, students, and administration as they work to support LMUs environmental sustainability endeavors. As a student involved in LMUs environmental sustainability and with many departments and individuals who work endless to advance LMUs ES efforts, I was greatly curious and intrigued with finding out why many of the universitys efforts failed to come to fruition. Over the years, I learned that across and within campus departments, information accessibility was limited. It seemed as though only a few key individuals knew information that would benefit and improve the work of others within the field.

Social scientists use Social Network Analysis metrics to understand how information travels amongst a group of people. These advanced metrics allow researchers to identify unique characteristics about the behavior of the group of people, the network. Centrality measurements utilize information about individuals network behavior and how they communication with others within the network to identify who within the group of people is most influential in spreading information. Banerjee et al. (2016) developed a new model that uses choice behavior, learning, and utility to find information diffusively central individuals. This Network Gossip model reforms the earlier

Diffusion Centrality model, and generalizes it so that future studies may apply the calculations without parameter restrictions.

Although the Network Gossip model does not require particular parameters, successful application of the model depends heavily on collecting the correct and appropriate data. Banerjee et al. (2016) use surveys with questions specifically designed to gather this correct data. Failure to replicate the meaning of the questions impacts the quality of the data collected. The appropriate data helps to build and characterize the informal, social network behavior of the group of people. Data calculations identify diffusively central individuals without first having to seed and track information. This makes the model much more adaptable and usable in various context. Thus, I used this new Network Gossip model to examine and test the communication behavior as it relates to LMUs environmental sustainability.

Q1. WHAT IF ANY, ARE THE DIFFERENCES BETWEEN THE FORMAL AND INFORMAL COMMUNICATION NETWORK STRUCUTRES ABOUT LMUs ENVIRONMENTAL SUSTAINABILITY?

Differences between the two networks can be found through comparing three different sources of data. First, the graphs provide a visual representation of network communication and nodal relationships, which allows us to draw basic conclusions about the networks global properties and local properties. Second, quantitative data gives use hard values for each nodes diffusion centrality. In understanding which node most effectively spreads information about a network, we can begin to understand with more depth the behavior characteristics and dynamics of the group of individuals. Third, qualitative, open-ended responses help to provide new insight into the results of the quantitative values, and they can support or call into question theses values.

Visual graphic properties are the easiest to compare, but provide a number of limitations to profound insight. The formal network illustrates how individuals, a designated by their job roles, tend to have nearly complete communication within their designated department. This is not surprising, however, since LMU, as a part of the universitys core values and general nature of the organization, supports and encourages an inclusive community . . . that is characterized by open dialogue, respect for individual differences, and collaboration across organizational boundaries [28]. The university has chosen to integrate their core values into its structure and operations through job responsibilities that require departmental and community collaboration. The formal network shows a great number of individuals within the network, suggesting that in fact a lot of individuals are involved in LMUs ES endeavors. The informal network, however, says otherwise. There are few individuals within the network and, particularly, a few larger nodes that suggest that even those few that are in the network believe that these large nodes take on a greater, more prominent role in LMUs environmental sustainability. Similarly, by looking at each nodes leading identification number and the links between the nodes, we see that relationship are distributed much more across departments than compact within a particular department. As such, we conclude that these key individuals work together and seek cross-departmental communication and information in order best do their job. These conclusions, however, are simply based on interpretation of graphic visuals, which may be biased based on the construction of the graph. Quantitative results, may add to these primary findings.

Quantitative results support the idea that a few key individuals dominate the communication within both the formal and informal networks. Individuals 211 and 116 are both listed in the top three most diffusively central individuals in both the formal and informal networks. Differences in those who are diffusively central, however, appear once we begin to compare the order of those beneath the top three. Individual 212, for instance, was actually replaced with individual 107 in the informal networks top three diffusively central individuals. Individuals 719 and 103 were also within the top five, however their diffusion centrality values drop down significantly from those listed above them. 116, ranked third, has a .4035 diffusion centrality value, 719 and 103 are almost half of that with values .2778 and .2390 respectively. 103s diffusion centrality in the formal network, however, is lower, and 719s diffusion centrality in the formal network is much lower.

Consideration for the data limitations from the Network Gossip model and my own adaptation of the models data collection methods, allow for a different analysis of the similarities and differences between the formal and informal networks quantitative data. If the informal network data picked up role expectation instead of actual communication behavior, then it should be expected that the top diffusively individuals match in both networksleaders and those with greater decision making authority should be the most diffusively central. The differences shown in the ranking of other individuals within the network may come from unclear information about what particular responsibilities each individual has in progressing LMUs ES. Qualitative responses explains, somewhat, the ambiguity with the quality of the quantitative data.

Open-ended responses allowed survey participants to express why they nominated certain individuals, or to explain more candidly their thought process behind particular nominations. (or lack there of). Many second survey respondents explained that they did not know which individuals fit the ES qualities and responsibility mentioned in the survey questions. A few offered nominations of individuals and departments though with an explanation that these nominations should theoretical taken on the characteristics and responsibilities mentioned in the questions, but that in actuality, they do not. These responses propose that the data collected and used to run the Diffusion Centrality calculations are tainted by expectation and chosen according to job title rather than actual network communication. As such, conclusion drawn from comparing the formal and informal network data can amount only to the conclusion that network knowledge about who is and should be involved with LMUs ES, and each individuals role within the network is imperfect.

Imperfect information and lack of access to this information decreases effective collaboration and integration of social, human, and physical capital. Similarly, if individuals within a network are unsure about the position and duties of those with whom they must work, the group loses its ability to hold individuals accountable. As a result, otherwise meaningful interactions between individuals may become inconsequential and knowledge share amongst the group reduces so that become much harder to make decisions and it takes much longer to accomplish new goals.

Q2. DO THE RELATIONSHIPS IDENTIFIED THROUGH THE NETWORK GOSSIP MODEL ACCURATELY DEPICT INDIVIDUALS RELATIONSHIP INSOFAR AS THE RELATE TO COMMUNICATION ABOUT LMUS ENVIRONMENTAL SUSTAINABILITY?

Over the course of my research, as I learned more about the methodology and mathematics behind the Network Gossip model, the more I became convinced that the model should accurately depict network communication behavior. The weakness in the construction of second surveys questions, caused impurities in my data, and thus left my quantitative results full of ambiguity and inclusiveness. Should research be able to perfect the composition of the survey questions so that the participants interpretation of the tone and purpose of the question are accurately communicated, I believe that research may even be able to reduce the effect of leadership status on the data. In other words, the correct questions may able to reduce the likelihood that individuals nominate others solely because of their leadership role or title within the network, but instead, because of true interaction and communication with that person.

Q3. WHAT IF ANYTHING DO THE CONSTRUCTED NETWORKS TELL ABOUT LMUS ENVIRONMENTAL SUSTAINABILITY EFFORTS AND LMU IN GENERAL?

As noted before, is it evident from both the formal and the informal networks that LMU values open communication and collaboration among individuals and across departments. Realistically, however, effective communication is not one of LMUs strong points. Similarly, they show that the university hosts a number of departments, faculty, staff, and administration who are involved and would like to be involved in advancing LMUs ES whether it be because of their specific job roles or due to personal interest, or both. LMU has the foundation upon which to excel and leader ES campus efforts, instead lack of accessible information restrict effective communication, heighten oversight of tasks and problems, and reduce important teamwork and collaboration.

Q4. HOW MIGHT THE IDENTIFIED WHOLE NETWORK PROPERTIES AND SPECIFIC NODAL PROPERTIES HELP INDIVIDUALS AND DEPARTMENTS ADVANCE THE UNIVERSITYS ENVIRONMENTAL SUSTAINABILITY EFFORTS?

Although LMU has the human capital, values, and support to progress ES endeavors, network analysis from this study suggests that individual lack important information, and thus, I assume some sort of barrier in communication of this important information. As suggested by the qualitative data, individuals involved in LMUs ES endeavor desire some sort of formal assembly to discussion decisions, tasks, plans, goals, innovations, and initiatives. Such a meeting would allow for more effective teamwork and operations and consequently, help further the universitys ES efforts and leadership.

Q5. WHAT ARE THE UNIVERSITY-LEVEL OUTCOMES OF THE CURRENT COMMUNICATION STRUCTURE?

Since the limitation in my data prevent me from speaking exactly about the actual ES communication network structure, it is understood that currently communication at LMU, whatever the structure may be, can be stronger. Without effect communication and teamwork, the university may be come a uncontrolled mess of ideas that never manifest into successful, sustainable action. Long-term success of environmental sustainability efforts falter and attention of university members gets drawn elsewhere. Individuals are eager to join forces to advance LMUs ES, but lack of appropriate knowledge prevents them from doing their best, superior work.

10.1 Future Research

Much of my results are inconclusive in identifying the de facto informal communication network for LMUs environmental sustainability, restructuring and redoing this study would deepen LMUs understanding of strengths and weakness of their operations and ES efforts. I would also like to seek ways in which to design the study to reduce the effect of status and title on individuals nominations. In doing so, I will help advance the metric and provide universities and all organizations with a valuable method through which they can understand and improve their operations. In understanding the informal network, administration can promote an environment that encourages and utilizes social capital for the overall success of the organization. Additionally, this study does not incorporate the effect students have on which decisions and which activities get the most attention. Thus, it will be interesting to extend the Network Gossip model to include students, their persistence with demands and interests, and their fleeting time and availability at the university.

APPENDIX

Appendix A

The image is an example of a Super Mario Dwarf character scaled up using bitmap, on the top, and vector art, on the bottom. The image demonstrates the effect of vector art and eigenvectors scalability capabilities. (Verbose, http://www.vg-resource.com/thread-26698-post-582314.html#pid582314)

Appendix B

Suyeon Khim, Perron-Frobenius Proof

THE FROBENIUS-PERRON THEOREM

SUYEON KHIM

1. Introduction

We begin by stating the Frobenius-Perron Theorem:

Theorem 1.1 (Frobenius-Perron)**.** *Let B be an $n \times n$ matrix with nonnegative real entries. Then we have the following:*

(1) *B has a nonnegative real eigenvalue. The largest such eigenvalue, $\lambda(B)$, dominates the absolute values of all other eigenvalues of B. The domination is strict if the entries of B are strictly positive.*
(2) *If B has strictly positive entries, then $\lambda(B)$ is a simple positive eigenvalue, and the corresponding eigenvector can be normalized to have strictly positive entries.*
(3) *If B has an eigenvector v with strictly positive entries, then the corresponding eigenvalue λ_v is $\lambda(B)$.*

We will first illustrate the statement for 2-by-2 matrices (using very elementary arguments), and then prove the theorem for the n-by-n case. Finally, we will conclude with examples of some of the applications of the theorem.

2. The Frobenius-Perron Theorem for $n = 2$

Consider the matrix

$$B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

with nonnegative entries. The characteristic polynomial

$$p_B(t) = \det(tI - B) = t^2 - (a+d)t + (ad - bc).$$

has discriminant

$$(a-d)^2 + 4bc \geq 0$$

and roots

$$\lambda(B) = \frac{(a+d) + \sqrt{(a-d)^2 + 4bc}}{2}, \quad \lambda'(B) = \frac{(a+d) - \sqrt{(a-d)^2 + 4bc}}{2}.$$

(1). Since $a, b, c, d \geq 0$, the discriminant is nonnegative, so the roots of the characteristic polynomial can only take on real values. Hence there exists a real eigenvalue for B. $\lambda(B)$ is nonnegative, so B has a nonnegative real eigenvalue. Since

$$t^2 - (a+d)t + (ad - bc) = \left[t - \frac{(a+d)}{2}\right]^2 - \left[\frac{(a-d)^2}{4} + (ad - bc)\right]$$

and $\frac{(a+d)}{2}$ is nonnegative, $\lambda(B) \geq |\lambda'(B)|$. If B has strictly positive entries, then $\frac{(a+d)}{2}$ is strictly positive and the domination is strict.

(2). If B has strictly positive entries, then the discriminant is greater than 0, so the characteristic polynomial must have two distinct real solutions. Of these, $\lambda(B)$ is positive and greater than $\lambda'(B)$. Hence, $\lambda(B)$ is a simple positive eigenvalue.

We now show that the eigenvector corresponding to $\lambda(B)$ can be normalized to have strictly positive entries. Define

$$D := (a-d)^2 + 4bc, \qquad \lambda := \lambda(B).$$

There exists an eigenvector x with eigenvalue λ. This eigenvector must be unique up to scaling, because there are two distinct eigenvalues, each with at least one corresponding eigenvector, and each with at most one corresponding eigenvector (up to scaling), since the number of linearly independent eigenvectors of a matrix cannot exceed its size. We have:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \end{pmatrix},$$

$$\begin{cases} ax_1 + bx_2 = \lambda x_1 \\ cx_1 + dx_2 = \lambda x_2 \end{cases}.$$

By definition, either $x_1 \neq 0$ or $x_2 \neq 0$. Suppose $x_1 \neq 0$. Then

$$\begin{aligned} a + b \cdot \frac{x_2}{x_1} = \lambda \qquad &\Leftrightarrow \qquad \frac{x_2}{x_1} = \frac{\lambda - a}{b}, \\ c + d \cdot \frac{x_2}{x_1} = \lambda \cdot \frac{x_2}{x_1} \qquad &\Leftrightarrow \qquad \frac{x_2}{x_1} \cdot (\lambda - d) = c > 0. \end{aligned}$$

We want to prove that $\frac{x_2}{x_1} > 0$. It is enough to show that either $\lambda > a$ or $\lambda > d$. This is indeed true, because $\lambda > \frac{a+d}{2}$. The same method proves the result for $x_2 \neq 0$.

(3). Suppose B has an eigenvector v with strictly positive entries. We have:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} \lambda_v v_1 \\ \lambda_v v_2 \end{pmatrix},$$

which we know from the proof of (2) gives us

$$a + b \cdot \frac{v_2}{v_1} = \lambda_v, \qquad c + d \cdot \frac{v_2}{v_1} = \lambda_v \cdot \frac{v_2}{v_1}.$$

From this, we obtain

$$\frac{v_2}{v_1} \cdot b = \lambda_v - a, \qquad \frac{v_2}{v_1} \cdot (\lambda_v - d) = c \geq 0.$$

Since $\frac{v_2}{v_1}$ is positive, we must have $\lambda_v \geq a$ and $\lambda_v \geq d$. Then $\lambda_v \geq \frac{a+d}{2}$, hence $\lambda_v = \lambda(B)$.

3. Proof of the Frobenius-Perron Theorem for n-by-n matrices

Now that we understand the theorem for $n = 2$, we will prove the general case. We will begin by proving (3), and furthermore show that if the entries of B are strictly positive, then the domination is strict. We will then show that λ_v is a simple positive eigenvalue, and the corresponding eigenvector can be normalized to have strictly positive entries. Next, we will show that the proof of (1) can be reduced to the case for B with strictly positive entries. Then by the above, it will suffice to prove the existence of an eigenvector v with strictly positive entries for B with strictly positive entries to conclude the proof of (1) and (2). We will prove the existence of such a v.

Proof of (3). Suppose B has an eigenvector v with strictly positive entries, and let λ_v denote the corresponding eigenvalue, so that $Bv = \lambda_v v$. Observe that

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} v_1 & 0 & \dots & 0 \\ 0 & v_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & v_n \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = C \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

where we denote the diagonal matrix by C in the last equality. Then

$$B \cdot C \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \lambda_v C \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \Longrightarrow C^{-1} B C \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \lambda_v \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Since

$$C^{-1} = \begin{pmatrix} v_1^{-1} & 0 & \dots & 0 \\ 0 & v_2^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & v_n^{-1} \end{pmatrix},$$

the matrix CBC^{-1} has only nonnegative entries. Similar matrices have the same eigenvalues, so we may assume without loss of generality that

$$v = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

We then have $\lambda_v = \sum_{j=1}^n b_{ij}$ for each $1 \leq i \leq n$. Hence λ_v is a nonnegative real number, and it is strictly positive unless $B = 0$.

Let us equip $\mathbb{C}^n$ with the ℓ^∞ norm, i.e.,

$$||z|| = \max_{i=1,\cdots,n} |z_i| \quad \text{for} \quad z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}.$$

For any $z \in \mathbb{C}^n$, the i-th entry of the vector Bz is equal to $b_{i1}z_1 + b_{i2}z_2 + \cdots + b_{in}z_n$. We have

$$|b_{i1}z_1 + \cdots + b_{in}z_n| \leq |b_{i1}||z_1| + \cdots + |b_{in}||z_n| \tag{3.1}$$

$$\leq \sum_{j=1}^n b_{ij} \cdot \max_{i=1,\ldots,n} |z_i| \tag{3.2}$$

$$= \lambda_v ||z||.$$

Therefore,

$$||Bz|| \leq \lambda_v ||z||.$$

Hence, if z' is an eigenvector with eigenvalue λ', then

$$||Bz'|| = |\lambda'| \cdot ||z'|| \leq \lambda_v ||z'||.$$

Therefore, $\lambda_v \geq |\lambda'|$. Hence, by definition, $\lambda_v = \lambda(B)$, as claimed.

Remark 1. Now suppose that all entries of B are strictly positive. Then $||Bz|| < \lambda_v ||z||$, unless $z_1 = z_2 = \cdots = z_n$, which is the same as saying

$$z = c \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = c \cdot v,$$

where $c \in \mathbb{C}$. This is the only case for which we have equality, hence v is the unique (up to scale) eigenvector with eigenvalue λ_v. This is because if $z_i \neq z_j$ for some $1 \leq i, j \leq n$, then one of the inequalities (3.1) or (3.2) will be strict. Then $||z||$ cancels on both sides ($||z||$ is greater than 0 by the definition of an eigenvector), and we see that λ_v strictly dominates the absolute values of all other eigenvalues of B. Hence, we have strict inequality for all eigenvalues corresponding to eigenvectors other than v.

Remark 2. We will now prove by contradiction that the "algebraic" multiplicity of λ_v (i.e., the multiplicity of λ_v as a root of the characteristic polynomial of B) is exactly 1. Suppose the multiplicity of λ_v is greater than 1. By the Jordan theorem, there exists an invertible matrix C such that CBC^{-1} is upper triangular and looks like the following matrix:

$$\begin{pmatrix} \ddots & & & 0 \\ & \lambda_v & 1 & \\ & & \lambda_v & \\ 0 & & & \ddots \end{pmatrix},$$

with a Jordan block of size at least 2. Note that we may exclude the case with two 1-by-1 Jordan blocks with the same λ_v, because then we would have two independent eigenvectors for λ_v, but we proved in Remark 1 that v is unique up to scalar multiple. We make the following claim:

Claim 1.

(i) There exist entries of $(\frac{1}{\lambda_v}CBC^{-1})^n$ such that the absolute values of these entries approach ∞ as $n \to \infty$.
(ii) Hence, the same is true for $(\frac{1}{\lambda_v}B)^n$.

Proof. (i)

$$\frac{1}{\lambda_v} \cdot \begin{pmatrix} \lambda_v & 1 & & 0 \\ & \lambda_v & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_v \end{pmatrix} = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix} + \begin{pmatrix} 0 & \frac{1}{\lambda_v} & & 0 \\ & 0 & \ddots & \\ & & \ddots & \frac{1}{\lambda_v} \\ 0 & & & 0 \end{pmatrix}.$$

Let

$$\begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix} = I, \qquad \begin{pmatrix} 0 & \frac{1}{\lambda_v} & & 0 \\ & 0 & \ddots & \\ & & \ddots & \frac{1}{\lambda_v} \\ 0 & & & 0 \end{pmatrix} = A.$$

We need only look at this particular Jordan block, because when we multiply the Jordan block decomposed matrix CBC^{-1} by itself, each Jordan block is only affected by its corresponding Jordan block. Furthermore, for $k > 1$, each A^k affects only its particular diagonal line of entries, from $a_{1(k+1)}$ to $a_{n(n-k)}$. By the binomial theorem, we have that

$$\begin{aligned}(I+A)^n &= \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} I^{n-k}A^k \\ &= \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} A^k \\ &= I + nA + \begin{pmatrix} n \\ 2 \end{pmatrix} A^2 + \cdots \\ &= \begin{pmatrix} 1 & \frac{n}{\lambda_v} & & * \\ & 1 & \ddots & \\ & & \ddots & \frac{n}{\lambda_v} \\ 0 & & & 1 \end{pmatrix}.\end{aligned}$$

This shows that $(\frac{1}{\lambda_v}CBC^{-1})^n$ has entries whose absolute values approach ∞ as $n \to \infty$.

(ii) To show that $||B^n|| \to \infty$ as $||(CBC^{-1})^n|| \to \infty$, note that $(CBC^{-1})^n = CB^nC^{-1}$. Think of these n-by-n matrices as elements of $\mathbb{C}^{n^2}$. Consider the function $f : \mathbb{C}^{n^2} \to \mathbb{C}^{n^2}$, where $f(X) = CXC^{-1}$. This function is continuous. Its inverse, $f^{-1} : \mathbb{C}^{n^2} \to \mathbb{C}^{n^2}$ is also continuous, where $f^{-1}(X) = C^{-1}XC$. Therefore, the entries of $(CBC^{-1})^n$ are bounded iff B^n is bounded.

$\square$

However, observe that the entries of $(\frac{1}{\lambda_v}B)^n$ cannot approach ∞, since

$$||Bz|| \le \lambda_v||z|| \quad \Leftrightarrow \quad \frac{1}{\lambda_v}||Bz|| = \left|\left|\frac{1}{\lambda_v}Bz\right|\right| \le ||z||,$$

and therefore

$$\left|\left|\left(\frac{1}{\lambda_v}B\right)^n z\right|\right| = \left|\left|\frac{1}{\lambda_v}\left[\left(\frac{1}{\lambda_v}\right)^{n-1}B^n\right]z\right|\right| \le ||z|| \tag{3.3}$$

for any $||z||$, since $(\frac{1}{\lambda_v})^{n-1}B^n$ is just another matrix with strictly positive entries and therefore can be substituted for B in the inequality (3.3). For any matrix A that has the property $||Az|| \le ||z||$, if $a \in \mathbb{C}^n$ is any row vector of A, then $||a \cdot z|| \le ||z||$. Then we have that the entries of A are bounded, since $|a_{ij}| \le \frac{2||z||}{|z_j|}$, Therefore, the entries of $(\frac{1}{\lambda_v}B)^n$ must be bounded. We have a contradiction, which shows that the algebraic multiplicity of λ_v cannot be greater than 1.

Proof of (1) and (2). We will reduce the proof of (1) to the case where all entries of B are strictly positive. The idea is that we may "approximate" B by matrices with strictly positive entries. Consider B with nonnegative entries. Define B_r to be the same matrix with the 0 entries replaced by $\frac{1}{r}$, where $r \in \mathbb{R}$ and $r > 0$. We will:

(i) show that the eigenvalues of B_r approach the eigenvalues of B as $r \to \infty$;
(ii) prove the existence of an eigenvector for B_r with strictly positive entries, and hence a positive eigenvalue for B_r, by our proof of (3) and the remarks; and
(iii) prove that this positive eigenvalue is precisely $\lambda_v = \lambda(B)$ and satisfies the properties stated in parts (1) and (2) of Theorem 1.1.

Since the eigenvalues of a matrix are the roots of its characteristic polynomial, if we show that as polynomials approach polynomials, roots approach roots, then we will have proved (i). We will use the following lemma:

Lemma 1. *Let* $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a^n, a_i \in \mathbb{C}$, $|a_1|, \cdots, |a_n| < M$. *If* $z \in \mathbb{C}$ *is a root of* f, *then* $|z| < 1 + nM$.

Proof. Suppose $|z| \geq 1 + nM$. If $f(x) = 0$, then

$$x^n = -a_{n-1}x^{n-1} - a_{n-2}x^{n-2} - \cdots - a_0.$$

Taking the absolute values of both sides,

$$\begin{aligned} |x^n| &= |-a_{n-1}x^{n-1} - a_{n-2}x^{n-2} - \cdots - a_0| \\ &= |a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_0|. \end{aligned}$$

We divide by x^n and obtain

$$\begin{aligned} 1 &= |a_{n-1}x^{n-1}/x + a_{n-2}x^{n-2}/x + \cdots + a_0/x| \\ &\leq \left|\frac{a_{n-1}}{x}\right| + \left|\frac{a_{n-2}}{x^2}\right| + \cdots + \left|\frac{a_0}{x_n}\right| \\ &< \left|\frac{M}{x}\right| + \left|\frac{M}{x^2}\right| + \cdots + \left|\frac{M}{x^n}\right| \\ &\leq \left|\frac{M}{1+nM}\right| + \left|\frac{M}{(1+nM)^2}\right| + \cdots + \left|\frac{M}{(1+nM)^n}\right| \\ &\leq n\left|\frac{M}{1+nM}\right| \\ &< 1. \end{aligned}$$

Contradiction.

□

This lemma establishes an upper bound for the absolute values of the roots. Arrange the eigenvalues of B_r in any order; call them $\lambda_1^{(r)}, \cdots, \lambda_n^{(r)}$. Since the sequence $\left\{\left(\lambda_1^{(r)}, \cdots, \lambda_n^{(r)}\right) \in \mathbb{C}^n\right\}_{r=1}^{\infty}$ is bounded, it has a convergent subsequence. Call it $\left\{\left(\lambda_1^{(r_j)}, \cdots, \lambda_n^{(r_j)}\right)\right\}_{r=1}^{\infty}$. Then

$$B_{r_j} \to B \quad \Longrightarrow \quad p(B_{r_j}) \to p(B)$$

as $j \to \infty$, where $p(B)$ denotes the characteristic polynomial of B. If we put

$$\lambda_k := \lim_{j \to \infty} \lambda_k^{(r_j)},$$

this implies that

$$p(B) = \prod_{k=1}^{n} (t - \lambda_k),$$

because

$$p(B_{r_j}) = \prod_{k=1}^{n} (t - \lambda_k^{(r_j)})$$

for every j. Therefore, there exists a subsequence such that the n-tuple of roots $\lambda_1^{(r)}, \cdots, \lambda_n^{(r)}$ converge to the n-tuple of roots of B (i.e., the eigenvalues). Hence, the λ_k's are the eigenvalues of B.

We will now prove the existence of an eigenvector with strictly positive entries for B with strictly positive entries. We will use the following claim:

Claim 2. *If $Bv' = \lambda_{v'} v'$ with the entries of v' being nonnegative, $v' \neq 0$, and the entries of B being strictly positive, then each entry of v' must be positive.*

Proof. Since all entries of v' are nonnegative, the same is true of Bv'. Furthermore, all entries of B are positive, so the entries of Bv' are all positive, since there is at least one nonnegative, non-zero entry in v'. However, v' is an eigenvector, so Bv' is a scalar of multiple v', which requires it to have zero entries in the same locations as v'. Hence, none of the entries of v' can be zero.

□

So we may prove the existence of an eigenvector v' with nonnegative entries for B with strictly positive entries, which by the claim is equivalent to proving the existence of an eigenvector v with strictly positive entries.

Let us consider the cube $D : \{d \in D \mid 0 \leq d_i \leq 1,\ \forall i = i, \cdots, n\}$. For the matrix B, we will write $||B|| = \max_{i=1,\ldots,n} \sum_{j=1}^{n} |b_{ij}|$. Consider the function $f(d) = ||dBd^{-1}||$, where we consider $d = (d_1, \cdots, d_n) \in D$ as a diagonal matrix

$$d = \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_n \end{pmatrix}.$$

Note that f is only defined in the "interior", D^{int}, of the cube D, where $d \in D^{int}$ if $d_i \neq 0$ for all $i = 1, \dots, n$. Moreover, f is clearly continuous on D^{int}.

For each $\epsilon > 0$, let us define a set D^ϵ by $d \in D^\epsilon$ iff $d_i \geq \epsilon$ for all i and $\sum_{i=1}^n d_i = 1$. Note that $D^\epsilon \subset D^{int}$ and is closed and bounded, hence compact. We claim that there exists $d \in D^{int}$ such that $f(d') \geq f(d)$ for all $d' \in D^{int}$.

Claim 3. *For any sufficiently small $\epsilon > 0$, there exists $d'' \in D^\epsilon$ such that $f(d') \geq f(d'')$ for all $d' \in D^{int}$.*

Proof. Without loss of generality, we can assume $\sum_{i=1}^n d'_i = 1$, since we can rescale the sum of the coordinates of the vector to equal 1. Assume d' does not lie in D^ϵ. (If it does, then since D^ϵ is compact, the function f achieves a minimum on D^ϵ and we are done.) Then for some i, we have $d'_i < \epsilon$ and $d'_1 + \cdots + d'_{i-1} + d'_{i+1} + \cdots + d'_n > 1 - \epsilon$. So there exists $j \neq i$ such that $d'_j > \frac{1-\epsilon}{n-1}$. Take the ji-th entry of $f(d') = ||d'Bd'^{-1}||$:

$$f(d') \geq d'_j {d'_i}^{-1} b_{ji} > \frac{(1-\epsilon)b_{ji}}{(n-1)\epsilon}.$$

If we take $\epsilon \to 0$, i.e., if one of the coordinates of $d \in D$ approaches 0, then $f(d) \to \infty$, so f achieves its smallest value on some $d'' \in D^\epsilon \subset D^{int}$. Therefore, $f(d) \geq f(d'')$, $\forall d \in D^{int}$.

□

Claim 4. *d' is an eigenvector for B with eigenvalue $f(d') = \lambda$.*

Proof. Replacing B by $d'Bd'^{-1}$, we may assume without loss of generality that $d' = (1, \cdots, 1)$, by the same line of reasoning as given in the proof of (3). We have

$$\max_i \sum_{j=1}^n b_{ij} = \lambda$$

$$\max_i \sum_{j=1}^n d_i {d_j}^{-1} b_{ij} \geq \lambda, \quad \text{such that } d_k > 0,\ \forall k. \tag{1.1}$$

Let $S = \{i \mid \sum_{j=1}^n b_{ij} < \lambda\}$. We only need to show that $S = \varnothing$. We will prove this by contradiction. By our condition, $\forall\ d_1, \cdots, d_n$, there exists $i \in \{1, \cdots, n\}$ such that $d_i \sum_{j=1}^n {d_j}^{-1} b_{ij} \geq \lambda$, and at the very least $\sum_{j=1}^n {d_j}^{-1} b_{ij} \geq \lambda$. So if we take $(d_1, \cdots, d_n)$ "very close" to $(1, \cdots, 1)$, then $i \notin S$. For instance, we could take

$$d_i = \begin{cases} 1-\epsilon & (i \notin S) \\ 1 & (i \in S). \end{cases}$$

If $S \neq \varnothing$ and $\epsilon > 0$ is sufficiently small then $\max_i \sum_{j=1}^n d_i {d_j}^{-1} b_{ij} < \lambda$, contrary to (1.1).

□

Since we have shown that an eigenvector v with strictly positive entries indeed exists for B with strictly positive entries, and we know it has the corresponding eigenvalue λ_v by (3), we have that $\lim_{r\to\infty} B_r = B$ has a nonnegative real eigenvalue. By our previous lemma, there exists a subsequence such that the eigenvalues of B_r converge to the eigenvalues of B. For every r, B_r has a positive real eigenvalue that strictly dominates all the other ones, by (3'). Call this eigenvalue $\lambda_1^{(r)}$, and arrange the other eigenvalues of B_r in any order; call them $\lambda_2^{(r)}, \cdots, \lambda_n^{(r)}$. We know the sequence $\{(\lambda_1^{(r)}, \cdots, \lambda_n^{(r)}) \in \mathbb{C}^n\}_{r=1}^{\infty}$ converges to $\{(\lambda_1^{(r_j)}, \cdots, \lambda_n^{(r_j)})\}_{r=1}^{\infty}$

We conclude that:

(1) $\lambda_1^{(r_j)} > 0$ by construction $\Rightarrow \lambda_1$ is real and nonnegative; and
(2) for any $2 \le k \le n$, $\lambda_1^{(r_j)} > |\lambda_k^{(r_j)}|$, $\forall j$. By passing to the limit as $j \to \infty$, we obtain $\lambda_1 \ge |\lambda_k|$, $\forall 2 \le k \le n$.

This completes the proof of Theorem 1.1.

4. Applications

The Frobenius-Perron theorem has a natural interpretation in the theory of Markov chains, which in turn has applications in population modeling and biophysics, to name but a few. We will illustrate a few of these.

Suppose we have any vector $u_0 = (x, 1-x)$, which we multiply over and over by the "transition matrix"

$$A = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix},$$

Then $u_1 = Au_0$, $u_2 = Au_1$, $\cdots$, $u_k = A^k u_0$. The claim is that the vectors $u_0, u_1, \ldots$ will approach a "steady state", i.e., multiplying A will eventually cease to change the vector. The limit state for this particular example is $u_\infty = (.6, .4)$. Observe that $Au_\infty = u_\infty$:

$$\begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix} \begin{bmatrix} .6 \\ .4 \end{bmatrix} = \begin{bmatrix} .6 \\ .4 \end{bmatrix}.$$

u_∞ is an eigenvector with eigenvalue 1, and this makes it steady. But what is significant is that the final outcome *does not depend on the starting vector*; for *any* u_0, $A^k u_0$ will *always* converge to $(.6, .4)$ as $k \to \infty$.

Having a steady state does not alone imply that all vectors u_0 lead to u_∞. For example,

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

has the steady state

$$B \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

but the starting vector $u_0 = (0,1)$ will give $u_1 = (0,2)$ and $u_2 = (0,4)$. B has $\lambda = 1$ but it also has $\lambda = 2$. Any $|\lambda| > 1$ means blowup.

The explanation for the phenomenon that for some matrices, all vectors u_0 lead to u_∞, forms the basis for the theory of Markov chains. There are two special properties of A that guarantee a steady state u_∞. These properties define what is called a Markov matrix, and A above is just one particular example:

1. Every entry of A is nonnegative.
2. Every column of A adds to 1.

Two facts are immediate for any Markov matrix A:

(i) multiplying a nonnegative u_0 by A produces a nonnegative $u_1 = Au_0$; and
(ii) if the components of u_0 add to 1, so do the components of $u_1 = Au_0$.

Statement (ii) follows from the fact that the components of u_0 add to 1 when $[1, \cdots, 1]\, u_0 = 1$. This is true for each column of A by Property 2. Then by matrix multiplication, it is true for Au_0:

$$[1, \cdots, 1]\, Au_0 = [1, \cdots, 1]\, u_0 = 1.$$

The same applies to $u_2 = Au_1$, $u_3 = Au_2$, etc. Hence, every vector $u_k = A^k u_0$ is nonnegative with components adding to 1. These are “probability vectors”. The limit u_∞ is also a probability vector, but first we must prove that a limit exists. We will show that $\lambda = 1$ is an eigenvalue of A and estimate the other eigenvalues.

Theorem 4.1. *If A is a positive Markov matrix, then $\lambda_1 = 1$ is larger than any other eigenvalue. The eigenvector x_1 is the steady state: $u_k = x_1 + c_2(\lambda_2)^k x_2 + \ldots + c_n(\lambda_n)^k x_n$ always approaches $u_\infty = x_1$.*

Every column of $A - I$ adds to $1 - 1 = 0$. The rows of $A - I$ add up to the zero row. Those rows are linearly dependent, so $A - I$ is singular. Its determinant is zero, hence $\lambda_1 = 1$ must be an eigenvalue of A. Strict domination, and hence uniqueness, follows from (2) of the Frobenius-Perron theorem. The other eigenvalues gradually disappear because $|\lambda| < 1$. The more steps we take, the closer we come to u_∞.

Example. The fraction of Illinois's wild raccoons in Chicago starts at $\frac{1}{50} = .2$. The fraction outside Chicago is .98. Every month 80% of raccoons in Chicago leave Chicago, while 20% of raccoons in Chicago remain in Chicago. Furthermore, 5% of raccoons outside Chicago arrive in Chicago, while 95% of raccoons outside of Chicago remain outside Chicago. Hence, the probability vector is multiplied by the Markov matrix

$$A = \begin{bmatrix} .80 & .05 \\ .20 & .95 \end{bmatrix},$$

which gives us

$$u_1 = Au_0 = A \begin{bmatrix} .02 \\ .98 \end{bmatrix} = \begin{bmatrix} .065 \\ .935 \end{bmatrix}.$$

In one month, the fraction of raccoons in Chicago is up to .065. What is the eventual outcome?

Since every column of A adds to 1, nothing is gained or lost - we are simply moving a fixed number of raccoons. The fractions add to 1 and the matrix A keeps them that way. We want to know how they are distributed after k time periods - which leads us to A^k.

Solution. To study the powers of A we diagonalize it:

$$|A - \lambda I| = \begin{vmatrix} .80 - \lambda & .05 \\ .20 & .95 - \lambda \end{vmatrix} = \lambda^2 - 1.75\lambda + .75 = (\lambda - 1)(\lambda - .75).$$

$$A \begin{bmatrix} .2 \\ .8 \end{bmatrix} = \begin{bmatrix} .2 \\ .8 \end{bmatrix}, \qquad A \begin{bmatrix} -1 \\ 1 \end{bmatrix} = .75 \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

We have eigenvalues $\lambda_1 = 1$ and $\lambda_2 = .75$ with corresponding eigenvectors $x_1 = (.2, .8)$ and $x_2 = (-1, 1)$. The eigenvectors are the columns of S, where S is the eigenvector matrix, $A^k = S\Lambda^k S^{-1}$. The starting vector u_0 is a combination of x_1 and x_2:

$$u_0 = \begin{bmatrix} .02 \\ .98 \end{bmatrix} = \begin{bmatrix} .2 \\ .8 \end{bmatrix} + .18 \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Now multiply by A to find u_1. The eigenvectors are multiplied by $\lambda_1 = 1$ and $\lambda_2 = .75$:

$$u_1 = 1 \begin{bmatrix} .2 \\ .8 \end{bmatrix} + (.75)(.18) \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$
$$u_k = A^k u_0 = \begin{bmatrix} .2 \\ .8 \end{bmatrix} + (.75)^k (.18) \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

The eigenvector x_1 with $\lambda = 1$ is the steady state u_∞. The other eigenvector x_2 gradually disappears because $|\lambda| < 1$. In the limit, $\frac{2}{10}$ of the raccoons are in Chicago and $\frac{8}{10}$ are outside.

Although we arrived at this particular conclusion using diagonalization, Jordan decomposition can be used to justify the statement for non-diagonalizable matrices. With a positive Markov matrix, the powers A^k approach the rank one matrix that has the steady state x_1 in every column.

It is of interest to biophysicists to derive approximate analytic expressions for the fraction of mutant proteins that fold stably to their native structure as a function of the number of amino acid substitutions, and estimate the asymptotic behavior of this fraction for a large number of amino acid substitutions. Using Markov chain approximation, it is possible to model how such a fraction decays.

5. Appendix

There is also an alternate proof of the existence of an eigenvector with strictly positive entries for B with strictly positive entries, which is faster than the one we gave in §3, but uses a rather nontrivial result, namely, Brauer's fixed point theorem, stated as Theorem 5.1 below. (Note that the proof we presented in §3 is much more elementary.)

Proof. Let us consider the subset $\Delta \subset \mathbb{R}^n$ defined by $z \in \Delta$ iff $z_i \geq 0$ for all $i = 1, \cdots, n$ and $\sum_{i=1}^n z_i = 1$. This is what is called an $(n-1)$-dimensional simplex (for $n = 2$, we get an interval, for $n = 3$, a triangle, and so on). Then, let us consider the map $\Phi : \Delta \to \Delta$, defined as follows:

$$\Phi(z) = \frac{Bz}{(Bz \cdot (1, 1, \cdots, 1))},$$

where $Bz \cdot (1, 1, \cdots, 1)$ denotes the dot product of the vectors Bz and $(1, 1, \cdots, 1)$ (i.e., the sum of the coordinates of the vector Bz). Clearly, Φ is a continuous map, so by Brauer's fixed point theorem, there exists $z \in \Delta$ such that $\Phi(z) = z$. Hence, $z = \frac{Bz}{Bz \cdot (1, \cdots, 1)} \Rightarrow Bz = (Bz \cdot (1, \cdots, 1))z$, so z is an eigenvector with nonnegative entries.

$\square$

Theorem 5.1 (Brauer's fixed point theorem). *Let $f : \Delta^n \to \Delta^n$ be a continuous map from an n-dimensional simplex to itself. Then, it has a fixed point (i.e., there exists $z \in \Delta^n$ such that $f(z) = z$).*

References

[1] Gilbert Strang. Introduction to Linear Algebra, 3rd Edition. Wellesley-Cambridge Press. 2003.

The proof of the Frobenius-Perron Theorem given in this paper is based on the presentation contained in unpublished notes of Vladimir Drinfeld distributed to the participants of the Geometric Langlands seminar in January 2006. The examples of Markov matrices were taken from the source named above.

This paper was written for the summer 2007 REU at the University of Chicago. Many thanks to Mitya Boyarchenko, Akaki Tikaradze, and Calvin Lin for their time and input, and to Changming and Younghee Kang for their generous funding.

Appendix C

Below are the MatLab calculation for the example Network Gossip Hearing matrix g, showing the probability that each individual is to hear from another individual at T=7 and q=3.1430. By summing across columns, we get each individuals Diffusion Centrality value. A,B,C,D,E,F,G,H,I,J,K,L,M,N,O = 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15 respectively.

Variables – sum_new_g

sum_new_g | t | DC | g | lambda

15x15 double

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.0833	1.3111	0.2684	0.0899	0.0213	1.0427	1.5414	1.0080	0.3576	0.0182	0.0037	0.0031	0.3546	0.2860	0.0717
2	1.3111	1.2041	0.3951	0.1539	0.0391	0.8090	1.6195	1.4596	0.5477	0.0312	0.0093	0.0079	0.5397	0.4246	0.1231
3	0.2684	0.3951	0.2134	0.0939	0.0258	0.1817	0.4312	0.7601	0.3062	0.0199	0.0069	0.0059	0.3003	0.2319	0.0743
4	0.0717	0.1234	0.0747	0.2947	0.4554	0.0487	0.1327	0.2528	0.5057	0.0561	0.1416	0.0154	0.1064	0.0858	0.0360
5	0.0189	0.0326	0.0209	0.4568	0.2689	0.0117	0.0343	0.0875	0.1740	0.1424	0.4035	0.0453	0.0475	0.0363	0.0564
6	1.0427	0.8090	0.1817	0.0601	0.0134	0.6273	1.1884	0.6996	0.2415	0.0113	0.0024	0.0020	0.2394	0.1927	0.0487
7	1.5414	1.6195	0.4312	0.1649	0.0411	1.1884	1.6372	1.6053	0.5947	0.0329	0.0096	0.0083	0.5865	0.4623	0.1324
8	1.0080	1.4596	0.7601	0.3320	0.1008	0.6996	1.6053	1.5777	1.0858	0.0813	0.0258	0.0199	1.0663	0.8344	0.2517
9	0.2860	0.4246	0.2319	0.5043	0.1674	0.1927	0.4623	0.8347	0.4592	0.0353	0.0519	0.0108	0.3271	0.2524	0.0855
10	0.0206	0.0374	0.0244	0.0230	0.0066	0.0130	0.0395	0.0935	0.0777	0.1276	0.0014	0.3586	0.1986	0.1516	0.4055
11	0.0068	0.0172	0.0128	0.1483	0.4049	0.0044	0.0179	0.0453	0.0749	0.4038	0.1272	0.1269	0.0739	0.0581	0.1427
12	0	0	0	0	0	0	0	0	0	0	0	0	0.4665	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0.4665	0	0
14	0.3576	0.5477	0.3062	0.2696	0.0843	0.2415	0.5947	1.0855	0.8363	0.1583	0.0230	0.0484	0.4438	0.4582	0.5004
15	0.0899	0.1536	0.0935	0.0843	0.0244	0.0601	0.1646	0.3306	0.2682	0.4069	0.0066	0.1276	0.1842	0.4991	0.2792

Appendix D

Below are the MatLab calculation values from taking the dot product of the Hearing matrix. These are the summed values from the Hearing matrix g at T=7 and q=3.1430. A,B,C,D,E,F,G,H,I,J,K,L,M,N,O = 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15 respectively.

	1
1	7.4609
2	8.6750
3	3.3149
4	2.4011
5	1.8369
6	5.3602
7	10.0556
8	10.9082
9	4.3262
10	1.5791
11	1.6652
12	0.4665
13	0.4665
14	5.9553
15	2.7728

Appendix E

Below is the MatLab code corresponding to matrix g's calculations in Appendix C and D:

```
clear all; clc;
t = 1:7; %change this for longer/shorter steps!!

% if you want to enter an excel matrix, just search the xlsread fuction

% g=xlsread('example matrix.xlsx','A1:D4');

g =[0 1 0 0 0 1 1 0 0 0 0 0 0 0 0;
1 0 0 0 0 0 1 1 0 0 0 0 0 0 0;
0 0 0 0 0 0 0 1 0 0 0 0 0 0 0;
0 0 0 0 1 0 0 0 1 0 0 0 0 0 0;
0 0 0 1 0 0 0 0 0 0 1 0 0 0 0;
1 0 0 0 0 0 1 0 0 0 0 0 0 0 0;
1 1 0 0 0 1 0 1 0 0 0 0 0 0 0;
0 1 1 0 0 0 1 0 1 0 0 0 1 1 0;
0 0 0 1 0 0 0 1 0 0 0 0 0 0 0;
0 0 0 0 0 0 0 0 0 0 0 1 0 0 1;
0 0 0 0 1 0 0 0 0 1 0 0 0 0 0;
0 0 0 0 0 0 0 0 0 0 0 0 1 0 0;
0 0 0 0 0 0 0 0 0 0 0 0 1 0 0;
0 0 0 0 0 0 0 1 1 0 0 0 0 0 1;
0 0 0 0 0 0 0 0 0 1 0 0 0 1 0];
g_old=g;

lambda=eig(g);

q=(1/max(lambda))/1; %change this if you change t!
for step=1:length(t)
        new_g(:,:,step)=q.*g_old;
        g_old=new_g(:,:,step);
        g_old=g_old*g;
end
sum_new_g=sum(new_g,3);
one=ones(length(g),1);
DC=sum_new_g*one;
```

Appendix F

These are the regression tables from Banerjee et al. (2016) that show the regression coefficient values for both Poisson and OLS regressions, and single and bivariate regressions.

TABLE C.1. Factors predicting nominations

Panel A: Poisson Regression	(1)	(2)	(3)	(4)	(5)
Diffusion Centrality	0.625***				
	(0.075)				
Degree Centrality		0.490***			
		(0.067)			
Eigenvector Centrality			0.614***		
			(0.084)		
Leader				0.950***	
				(0.271)	
Geographic Centrality					-0.113
					(0.082)
Observations	6,466	6,466	6,466	5,733	6,466
Panel B: OLS	(1)	(2)	(3)	(4)	(5)
Diffusion Centrality	0.391***				
	(0.071)				
Degree Centrality		0.367***			
		(0.065)			
Eigenvector Centrality			0.378***		
			(0.074)		
Leader				0.629***	
				(0.229)	
Geographic Centrality					-0.045
					(0.029)
Observations	6,466	6,466	6,466	5,733	6,466

TABLE C.2. Factors predicting nominations

Panel A: Poisson Regression	(1)	(2)	(3)	(4)	(5)
Diffusion Centrality	0.560***	0.431***	0.565***	0.624***	0.560***
	(0.122)	(0.130)	(0.086)	(0.075)	(0.122)
Degree Centrality	0.070				0.070
	(0.086)				(0.086)
Eigenvector Centrality		0.219			
		(0.138)			
Leader			0.613**		
			(0.290)		
Geographic Centrality				-0.115	
				(0.089)	
Observations	6,466	6,466	5,733	6,466	6,466
Post-LASSO					✓
Panel B: OLS	(1)	(2)	(3)	(4)	(5)
Diffusion Centrality	0.310***	0.266***	0.383***	0.391***	0.310***
	(0.112)	(0.089)	(0.081)	(0.071)	(0.112)
Degree Centrality	0.091				0.091
	(0.079)				(0.079)
Eigenvector Centrality		0.138			
		(0.089)			
Leader			0.457*		
			(0.231)		
Geographic Centrality				-0.045	
				(0.030)	
Observations	6,466	6,466	5,733	6,466	6,466
Post-LASSO					✓

Appendix G

The following is the first survey that I sent out to 41 individuals at the beginning of the study. I sent the survey by email and hardcopy via LMU's internal mail, and received a 61% response rate.

On-Campus Environmental Sustainability Involvement Questionnaire

Purpose of the Questionnaire:

To identify and build a visually representative network for communication and spread of environmental sustainability information, which affects that growth and implementation of environmental sustainability programs and practices on LMU's campus.

Definitions used for the following study:

Environmental Sustainability[1] – This definition is a three-fold definition taken from Herman Daly: 1. Renewable resources, the rate of harvest should not exceed the rate of regeneration; 2. Pollution, the rates of waste generation from projects should not exceed the assimilative capacity of the environment; and 3. Nonrenewable resources, the depletion of the nonrenewable resources should require comparable development of renewable substitute for that resource.

Involvement – Participation in the advancement of environmental sustainability operations and practices on LMU's campus, and all activity that helps to reduce the campus' carbon footprint.

Survey

1. What is your role at LMU?

   ____
   ____

2. To whom to do you report directly?

   ____
   ____

[1]Daly, H. E. 1990a. Boundless bull. Gannett Center Journal 4(3):113–118. —Daly, H. E. 1990b. *Toward some operational*

3. Does your boss have a superior? If so, who?

____
____

4. What are your areas of focus as they relate to environmental sustainability at LMU?

____
____

____
____

5. What types of environmental sustainability projects are you involved in at LMU?

____
____

6. For what sort of items (related to the reduction of LMU's campus carbon footprint) related to environmental sustainability on campus do you have the authority to approve?

____
____

7. To what extent are you involved in the development (identification and elaboration) of environmental sustainability operations on campus?
   1 2 3 4 5 6
   7
   (Not at all) (Somewhat) (Very
   much so)

8. To what extent are you involved in the implementation (generation, evolution, and carry out) of environmental sustainability operations on campus?
   1 2 3 4 5 6
   7
   (Not at all) (Somewhat) (Very
   much so)

9. To what extent are you involved in the continued management (on-going control, progress, and function) of environmental sustainability operations on campus?
   1 2 3 4 5 6
   7

(Not at all) (Somewhat) (Very much so)

10. To what extent are you involved in the education of environmental sustainability on campus?

1 2 3 4 5 6 7

(Not at all) (Somewhat) (Very much so)

Explain:____

____

____

____

11. How many hours a day do you spending working in total?

1-4hr(s) 5-8hrs 9-12hrs 13-16hrs 17-20hrs 21-24hrs

12. How much of your time do you typically spend per day on the topic of environmental sustainability? Circle One:

1hr 2hrs 3hrs 4hrs 5hrs 6hrs 7hrs 8hrs 9hrs 10hrs 11hrs 12hrs+

If necessary,
Explain:____

____

____

13. What is your role in planning, decisions making, and achieving environmental sustainability on LMU's campus?

____
____

14. How many years have you worked/been at LMU?

____
____________yr(s)_____

15. Do you feel you have experience with environmental sustainability beyond that which you utilize here at LMU? Choose One: Yes No

    Explain:____
    ____

    ____
    ____

    ____
    ____

    ____
    ____

    ____
    ____

    ____
    ____

Other Comments/Suggestions:
____

____
____

____
____

____
____

____
____

Would you be interested in participating in a follow up in-person interview should we derive additional questions from your responses? Choose One: Yes No

Appendix H

The following is the second survey that I sent out to 23 individuals at the beginning of the study. I created the survey on Qualtrics and sent the survey link by email. I received a 65% response rate.

Coco Freling - Senior Thesis Survey

Q1 Hello. Thank you again for agreeing to participate and for taking the time to answer some further questions. For this study the following terms are defined as such: Environmental Sustainability – This three-fold definition is taken from Herman Daly: 1. Renewable resources: The rate of harvest should not exceed the rate of regeneration; 2. Pollution: The rate of waste generation from projects should not exceed the assimilative capacity of the environment; and 3. Nonrenewable resources: The depletion of the nonrenewable resources should require comparable development of renewable substitute for that resource. Involvement – Participation in the advancement of environmental sustainability operations and practices on LMU’s campus, and all activities that help to reduce the campus’ carbon footprint.

Q33 Please enter your name:

Q2 In the following questions, I will ask you to think about individuals and department who are influential in LMU's environmental sustainability efforts here on campus. If you identify as such an individual in any of the following questions, please feel free to mention your own name.

Q3 When looking for information regarding LMU's environmental sustainability, who are 3 individuals you would approach first? If possible, please list in rank order from 1 to 3.

Individual 1 (1)
Individual 2 (2)
Individual 3 (3)

Answer If Q3 Individual 1 Is Not Empty

Q4 To what extent do you believe ${q://QID3/ChoiceTextEntryValue/1} is involved in LMU's environmental sustainability efforts?

	Not At All Involved (1)	Slightly Involved (2)	Somewhat Involved (3)	Very Involved (4)	Extremely Involved (5)
${q://QID3/ChoiceTextEntryValue/1} (1)	❍	❍	❍	❍	❍

Answer If Q3 Individual 1 Is Not Empty

Q5 In which of the following areas of environmental sustainability at LMU is ${q://QID3/ChoiceTextEntryValue/1} most involved? Check all that apply.

- ❑ Water (1)
- ❑ CO2 Emission (2)
- ❑ Recycling (3)
- ❑ Food and Beverage (4)
- ❑ Compost (5)
- ❑ Campus Purchases (Office Supplies, etc.) (6)
- ❑ Electricity (7)
- ❑ Green House Gases (8)
- ❑ Campus Fleet; Fuel (9)
- ❑ Architecture and Building (10)
- ❑ Waste (11)
- ❑ Landscape (12)
- ❑ Other (Specify) (13)
- ❑ Other (Specify) (14)
- ❑ Other (Specify) (15)
- ❑ If necessary, please explain further: (16) ____________________

Answer If Q3 Individual 2 Is Not Empty

Q6 To what extent do you believe ${q://QID3/ChoiceTextEntryValue/2} is involved in LMU's environmental sustainability efforts?

	Not At All Involved (1)	Slightly Involved (2)	Somewhat Involved (3)	Very Involved (4)	Extremely Involved (5)
${q://QID3/ChoiceTextEntryValue/2} (1)	❍	❍	❍	❍	❍

Answer If Q3 Individual 2 Is Not Empty

Q7 In which of the following areas of environmental sustainability at LMU is ${q://QID3/ChoiceTextEntryValue/2} most involved? Check all that apply.

- ❑ Water (1)
- ❑ CO_2 Emission (2)
- ❑ Recycling (3)
- ❑ Food and Beverage (4)
- ❑ Compost (5)
- ❑ Campus Purchases (Office Supplies, etc.) (6)
- ❑ Electricity (7)
- ❑ Green House Gases (8)
- ❑ Campus Fleet; Fuel (9)
- ❑ Architecture and Building (10)
- ❑ Waste (11)
- ❑ Landscape (12)
- ❑ Other (Specify) (13)
- ❑ Other (Specify) (14)
- ❑ Other (Specify) (15)
- ❑ If necessary, please explain further: (16) ____________________

Answer If Q3 Individual 3 Is Not Empty

Q8 To what extent do you believe ${q://QID3/ChoiceTextEntryValue/3} is involved in LMU's environmental sustainability efforts?

	Not At All Involved (1)	Slightly Involved (2)	Somewhat Involved (3)	Very Involved (4)	Extremely Involved (5)
${q://QID3/ChoiceTextEntryValue/3} (1)	❍	❍	❍	❍	❍

Answer If Q3 Individual 3 Is Not Empty

Q9 In which of the following areas of environmental sustainability at LMU is ${q://QID3/ChoiceTextEntryValue/3} most involved? Check all that apply.

- ❑ Water (1)
- ❑ CO_2 Emission (2)
- ❑ Recycling (3)
- ❑ Food and Beverage (4)
- ❑ Compost (5)
- ❑ Campus Purchases (Office Supplies, etc.) (6)
- ❑ Electricity (7)
- ❑ Green House Gases (8)
- ❑ Campus Fleet; Fuel (9)
- ❑ Architecture and Building (10)
- ❑ Waste (11)
- ❑ Landscape (12)
- ❑ Other (Specify) (13)
- ❑ Other (Specify) (14)
- ❑ Other (Specify) (15)
- ❑ If necessary, please explain further: (16) ____________________

Q10 For which topics related to LMU's environmental sustainability do you seek the most information? Check all that apply.

- ❑ Water (1)
- ❑ CO_2 Emission (2)
- ❑ Recycling (3)
- ❑ Food and Beverage (4)
- ❑ Compost (5)
- ❑ Campus Purchases (Office Supplies, etc.) (6)
- ❑ Electricity (7)
- ❑ Green House Gases (8)
- ❑ Campus Fleet; Fuel (9)
- ❑ Architecture and Building (10)
- ❑ Waste (11)
- ❑ Landscape (12)
- ❑ Other (Specify) (13)
- ❑ Other (Specify) (14)
- ❑ Other (Specify) (15)
- ❑ If necessary, please explain further: (16) ____________________

Q11 Nominate the top 3 individuals you recommend working with for the implementation of environmental sustainability practices at LMU.

Individual 1 (1)
Individual 2 (2)
Individual 3 (3)

Q12 Nominate 3 individuals that are influential in disseminating information around LMU regarding LMU's environmental sustainability efforts.

Individual 1 (1)
Individual 2 (2)
Individual 3 (3)

Q13 When you think of LMU's environmental sustainability efforts, which 3 individuals do you consider to be leaders in these efforts?

Individual 1 (1)
Individual 2 (2)
Individual 3 (3)

Q14 Which of the following areas of LMU's environmental sustainability do you think receive the greatest attention around campus? Check all that apply.

- ❑ Water (1)
- ❑ CO2 Emission (2)
- ❑ Recycling (3)
- ❑ Food and Beverage (4)
- ❑ Compost (5)
- ❑ Campus Purchases (Office Supplies, etc.) (6)
- ❑ Electricity (7)
- ❑ Green House Gases (8)
- ❑ Campus Fleet; Fuel (9)
- ❑ Architecture and Building (10)
- ❑ Waste (11)
- ❑ Landscape (12)
- ❑ Other (Specify) (13)
- ❑ Other (Specify) (14)
- ❑ Other (Specify) (15)
- ❑ If necessary, please explain further: (16) ____________________

Q15 Which LMU department plays a fundamental role in advancing environmental sustainability practices on campus? (i.e. without the support from this department environmental sustainability initiatives would lack the necessary evolution, innovation, and progression to keep up with federal and state requirements and overall social expectations). Please limit your response to your top answer.

Answer If Q15 Text Response Is Not Empty

Q16 In what way(s) does ${q://QID22/ChoiceTextEntryValue} help to advance LMU's environmental sustainability practices on campus?

Answer If Q15 Text Response Is Not Empty

Q17 How influential is ${q://QID22/ChoiceTextEntryValue} in progressing LMU's environmental sustainability efforts?

	Not At All Influential (1)	Slightly Influential (2)	Somewhat Influential (3)	Very Influential (4)	Extremely Influential (5)
${q://QID22/ChoiceTextEntryValue} (1)	❍	❍	❍	❍	❍

Q18 Which LMU department contributes most to the education about LMU's environmental sustainability? (i.e. this department strongly influences a change in LMU's individuals' behavior). Please limit your response to your top answer.

Answer If Q18 Text Response Is Not Empty

Q19 In what way(s) does ${q://QID25/ChoiceTextEntryValue} contribute to the education about LMU's environmental sustainability practices?

Answer If Q18 Text Response Is Not Empty

Q20 How influential is ${q://QID25/ChoiceTextEntryValue} in teaching about LMU's environmental sustainability efforts?

	Not At All Influential (1)	Slightly Influential (2)	Somewhat Influential (3)	Very Influential (4)	Extremely Influential (5)
${q://QID25/ChoiceTextEntryValue} (1)	❍	❍	❍	❍	❍

Q21 Which department at LMU plays a fundamental role in overseeing and preserving environmental sustainability practices on campus? (i.e. this department ensures that LMU's environmental sustainability initiative follow through and have long-term success). Please limit your response to your top answer.

Answer If Q21 Text Response Is Not Empty

Q22 In what way(s) does ${q://QID28/ChoiceTextEntryValue} carry out their overseeing and preserving roles?

Answer If Q21 Text Response Is Not Empty

Q23 Of the following areas of LMU's environmental sustainability, on which does ${q://QID28/ChoiceTextEntryValue} focus most of their time?

- ❑ Water (1)
- ❑ CO2 Emission (2)
- ❑ Recycling (3)
- ❑ Food and Beverage (4)
- ❑ Compost (5)
- ❑ Campus Purchases (Office Supplies, etc.) (6)
- ❑ Electricity (7)
- ❑ Green House Gases (8)
- ❑ Campus Fleet; Fuel (9)
- ❑ Architecture and Building (10)
- ❑ Waste (11)
- ❑ Landscape (12)
- ❑ Other (Specify) (13)
- ❑ Other (Specify) (14)
- ❑ Other (Specify) (15)
- ❑ If necessary, please explain further: (16) ____________________

Answer If Q21 Text Response Is Not Empty

Q24 How influential is ${q://QID28/ChoiceTextEntryValue} in the preservation of LMU's environmental sustainability efforts?

	Not At All Influential (1)	Slightly Influential (2)	Somewhat Influential (3)	Very Influential (4)	Extremely Influential (5)
${q://QID28/ChoiceTextEntryValue} (1)	❍	❍	❍	❍	❍

Appendix I

Below is the code that I used to find Diffusion Centrality. This code contains elements taken written by me with the help of Bryce Currey and taken from MIT's "Overview of metrics and their correlation patterns for multiple-metric topology analysis on heterogeneous graph ensembles. All codes listed below in the comment "This requires. . . " are MIT's.

```
%This requires diameter.m, dijkstra.m, adj2adjL.m, purge.m. retrieved from
http://strategic.mit.edu/downloads.php?page=matlab_networks


clear all; clc;
g=xlsread('/Users/afreling/Desktop/Cocos
Project/UncodedFormalNetwork.xlsx','A1:BF58'); %change according to document
path for the matrix.
d=diameter(g);
t=1:d;

%  g =[0 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0;
%     1 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0;
%     0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0;
%     0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0;
%     0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0;
%     1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0;
%     1 1 0 0 0 0 1 0 1 0 0 0 0 0 0 0;
%     0 1 1 0 0 0 1 0 1 0 0 0 1 1 0;
%     0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0;
%     0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1;
%     0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0;
%     0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0;
%     0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0;
%     0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 1;
%     0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0];

g_old=g;
lambda=eig(g);
q=(1/max(lambda))/1; %change this if you change t!
for step=1:length(t)
   new_g(:,:,step)=q.*g_old;
   g_old=new_g(:,:,step);
   g_old=g_old*g;
end
```

```
sum_new_g=sum(new_g,3);
one=ones(length(g),1);
DC=sum_new_g*one;
```

Code for Diameter.m:

```
function diam = diameter(adj)
diam = 0;
for i = 1:size(adj,1)
   [d,p] = dijkstra(adj,i,[]);
   for j=1:length(p);
      diam(i,j)=length(cell2mat(p(j)));
   end
end
diam=max(max(diam));
```

Code for dijkstra.m:

```
INPUTS: adj - adjacency matrix, s - source node, target - target node
% OUTPUTS: distance, d and path, P (from s to target)
% Note: if target==[], then dist and P include all distances and paths from s
% Other routines used: adj2adjL.m, purge.m
% GB, Last Updated: Dec 22, 2009

function [dist,P]=dijkstra(adj,s,target)

% initialize distances ========================
n=length(adj);          % number of nodes
adjL=adj2adjL(adj);      % list of neighbors

dist=inf(1,n);
dist(s)=0;

previous=[1:n; inf(1,n)]';  % {i: inf}, i=1:n, inf -> not assigned
S=cell(1,n); % shortest path sequence

Q=[1:n]; % all unvisited vertices, entire graph
while length(Q)>0 % while not empty
   % get min dist member among unvisited vertices
   [mindist,min_ind]=min(dist(Q))
   u=Q(min_ind);
```

```
    % termination condition - save source-u path
    S{u}=[];
    t=u;
    while not(isempty(find(previous(:,1)==t)))  % t in previous.keys():
       % insert u at the beginning of S
       S{u}=[t S{u}];
       t=previous(t,2);
    end

    if length(target)>0 & u==target
       dist=dist(u); P=S{u};
       return
    end

    % =========================================
    Q=purge(Q,u);  % remove u from Q
    for v=1:length(adjL{u})   % across all neighbors of u
       v=adjL{u}(v);
       alt=dist(u)+adj(u,v);
       if alt < dist(v)
          dist(v)=alt;
          previous(v,2)=u;
       end
    end
end
P=S;
```

Code for adj2adjL.m:

```
% Converts an adjacency graph representation to an adjacency list
% Valid for a general (directed, not simple) network model, but edge
% weights get lost in the conversion.
% INPUT: an adjacency matrix, NxN, N - # of nodes
% OUTPUT: cell structure for adjacency list: x{i_1}=[j_1,j_2 ...]
% GB, October 1, 2009

function L = adj2adjL(adj)

L=cell(length(adj),1);

for i=1:length(adj); L{i}=find(adj(i,:)>0); end
```

Code for purge.m:

```
% Removes a subset from a set, but preserves order of elements
% Similar to setdiff - which sorts the elements
% INPUTs: original set A, subset to remove B
% OUTPUTs: set Anew = A-B
% GB, Last updated: October 12, 2009

function Anew = purge(A,B)

Anew = [];
for a=1:numel(A);
  if isempty(find(B==A(a))); Anew=[Anew, A(a)]; end
end
```

Appendix J

The following is MIT's code specifically written to calculate Eigenvector centrality. This code is retrieved from
http://strategic.mit.edu/downloads.php?page=matlab_networks

```
% The ith component of the eigenvector corresponding to the greatest
% eigenvalue gives the centrality score of the ith node in the network.
% INPUTs: adjacency matrix
% OUTPUTs: eigen(-centrality) vector
% GB, Last Updated: October 14, 2009

function x=eigencentrality(adj)

[V,D]=eig(adj);
[max_eig,ind]=max(diag(D));
x=V(:,ind);
```

Bibliography

[1] Esther Duflo Abhijit Banerjee, Arun G. Chandrasekhar and Matthew O. Jackson. The diffusion of microfinance. *Science Magazine*, 341(6144), 2013.

[2] Esther Duflo Abhijit Banerjee, Arun G. Chandrasekhar and Matthew O. Jackson. The diffusion of microfinance. 2016.

[3] Daron Acemoglu and Asu Ozdaglar. 6/207/14.15: Networks, lecture 3: Erds-reyni graphs and branching processes. Technical report, Massachusetts Institute of Technology, http://economics.mit.edu/files/4621, 9 2009.

[4] Silvana Stefani Ahmad K. Naimzada and Anna Torriero. *Networks, Topology, and Dynamics.* Springer Science & Business Media, New York, New York, 2008.

[5] Phillip Bonacich. Power and centrality: A family of measures. *American Journal of Sociology.*, 92(1170-1182), 1987.

[6] G. Bounova and O.L. de Weck. Matlab tools for network analysis. Technical report, Massachusetts Institute of Technology, `http://strategic.mit.edu/downloads.php?page=matlab_networks`, 2011.

[7] Bratton. Networks: Basic concepts. Technical report, Louisiana State University, 2016.

[8] John Seely Brown and Paul Duguid. Organizing knowledge. *California Management Review*, 40(90-1111), 1998.

[9] Dr. Myla Bui. lecture.

[10] ClockBackward. Ordinary least squares linear regression: Flaws, problems, and pitfalls. Blog Essay, 6 2009.

[11] Michael D. Cohen, James G. March, and Johan P. Olsen. A garbage can model of organizational choice. *Administrative Science Quarterly*, (1-25), 1972.

[12] U.S. Green Building Council. Leedv4 building design & construction guide.

[13] Richard M. Cyert and James G. March. *The Tipping Point: How Little Things can Make a Big Difference.* Brown Little, Boston, Massachusetts, 2000.

[14] Mark Rennaker Dan Novak and Paulette Turner. Using organizational network analysis to improve across organizational boundaries. Technical report, YMCDN People & Strategy, `http://c.ymcdn.com/sites/www.hrps.org/resource/dynamic/blogs/20120224_111638_18645.pdf`, 2011.

[15] Merriam Webster Dictionary. "network".

[16] Oxford Dictionary. "network".

[17] Dictionary.com. "network".

[18] Stephen Dovers. Institutions for sustainability. Technical report, Centre for Resources and Environmental Studies, The Australian National University, https://digitalcollections.anu.edu.au/bitstream/1885/40972/3/een0101.pdf, 4 2001.

[19] EcoJesuit. Jesuit response to environmental and ecology challenges: Healing a broken world. *Ecology and Jesuits in Communication.*, 2011.

[20] The Association for the Advancement of Sustainability in Higher Education. About aashe.

[21] Eric Goodnight. What can super mario teach us about graphics technology?

[22] Robert A. Hanneman and Mark Riddle. *Introduction of Social Network Methods*, chapter 1, pages 1–18. Department of Sociology, The University of California, Riverside, `http://faculty.ucr.edu/~hanneman/nettext/Introduction_to_Social_Network_Methods.pdf`, 3 2005.

[23] Dr. Andrew Healy. lecture.

[24] Student Housing. Sustainability.

[25] Matthew O. Jackson. *Social and Economic Networks.* Princeton University Press, Princeton, New Jersey, 2008.

[26] Bjorn Jamtveit Joachim Mathieson and Kim Sneppen. Organization structure and communication networks in a university environment. *Physical Review E(Statistical, Nonlinear, and Soft Matter Physics).*, 82, 2010.

[27] Suyeon Khim. *The Frobenius-Perron Theorem.* PhD thesis, The University of Chicago, http://www.math.uchicago.edu/ may/VIGRE/VIGRE2007/REUPapers/FINALAPP/Khim.p 2007.

[28] LMU. Core values.

[29] Hilton School of Business Administration LMU Management. Lmu management.

[30] Henry Longbottom. Laudato si - green jesuit's 2 page summary. *Faith Seeking Environmental Justice.*, 2015.

[31] Drew Masters. Random graphs.

[32] John B. Miner. Organizational behavior 3: Historical origins, theoretical foundations, and the future. *M.E. Sharpe, Business & Economics*, pages 61–98, 2016.

[33] M.E.J. Newman. A measure of betweenness centrality based on random walks. Master's thesis, University of Michigan, https://arxiv.org/pdf/cond-mat/0309045.pdf, 9 2003.

[34] The City of Los Angeles and Impact Sciences. Draft environmental impact report. Technical report, Loyola Marymount University, `http://www.lmu.edu/Assets/University+Relations+Division/Communications+$!2b+Government+Relations/LMU_MP_DEIR_r3.pdf`, 1 2010.

[35] Office of the Federal Register. *What is the Electoral College.* National Archives and Records Administration, `http://www.archives.gov/federal-register/electoral-college/about.html`.

[36] Tore Opsahl. Closeness centrality in networks with disconnected components @blog, 3 2010.

[37] Piotr Brdka Przemyslaw Rewski, Jaroslaw Jankowski and Radoslaw Michalski. Knowledge workers collaborative learning behavior modeling in an organizational social network. *Computers in Human Behavior*, 51(1248-1260), 2015.

[38] Garry Robins. *Doing Social Network Research: Network-based Research Design for Social Scientists.* SAGE Publications, Thousand Oaks, California, 2015.

[39] Gideon Rosenblatt. What is network density - and how do you calculate it?

[40] Katelyn Samson Sarah A. Bekessy and R.E. Clarkson. The failure of non-binding declarations to achieve university sustainability: A need for accountability. *International Journal of Sustainability in Higher Education*, 8(301-316), 2007.

[41] John Scott. *Social Network Analysis.* SAGE Publishing, Thousand Oaks , California, 2013.

[42] Nist Sematech. 1.3.6.6.16. extreme value type i distribution.

[43] Leo Spizzirri. *Justification and Application of Eigenvector Centrality.* PhD thesis, Washington, `https://www.math.washington.edu/~morrow/336_11/papers/leo.pdf`, 3 2011.

[44] William Stonecypher. personal communication.

[45] and M.K. Holt T.M. Huning, P.C. Bryant. Informal social networks in organizations: Propositions regarding their role in organizational behavior outcomes. *Academy of Strategic Management Journal*, 14(20-29), 2014.

[46] Nikolas Tsakas. On decay centrality. *CoRR*, abs/1604.05582, 2016.

[47] North Carolina State University. Parameter identification of radioactive decay.