{Title}

Stephanie K. Ananth

Carnegie Mellon University

**Abstract**

Because I thought {X = Connection to themes of our class this semester. Do I understand the issues and questions at play?}, I did {Y = What I did. What domain? Is it a good approach? Will it actually get at the theme X?}, using method {Z = How well did I deploy the method? Do I understand what it can and can’t do? Did I implement it correctly? Do I consider alternate explanations, confounds, and how it might not have captured the full story?}. We found {A = What did I find? Am I at the right level of detail that that I the reader can follow my results without getting lost in irrelevant pieces?}. It means {B = What does A tell you about X? How does it deepen our understanding of the way the world works?}.

*Keywords:* page-rank, social prestige, social power, knowledge, distributed groups

**Theme**

In this paper, I will be exploring the applications of page-rank (or social rank/hierarchy) and social power/prestige in small-scale distributed groups over time. I will be utilizing data I collected almost exactly two years ago about distributed groups, and I will be looking at the following five dimensions:

* familiarity (how well one knows/is known by others in the group)
* favor (how much someone likes/is liked by others in the group)
* communication (how often communication happens [in both directions] between members of the group)
* trust (how much someone trusts/is trusted by others in the group)
* support (likelihood of asking/being asked by others in the group for advice or emotional support)

Just as hens abide by a defined “pecking-order”, humans abide by learned rules of social behavior. Social rank and hierarchy are aspects of every human community. Rarely written down or documented, these social guidelines are passed along through communities as tacit knowledge or as unspoken rules.

My goal is to define an explicit “pecking-order” for a small, distributed, former high school friend group comprised of 7 males and 7 females, all among the ages of 21-22 and to analyze possible interactions, connections, and explanations between each of the five dimensions I listed.

**Issues & Questions**

The issues and questions in play are primarily related to social dynamics and subjective, tacit knowledge. (The challenge is to draw out this knowledge without introducing biases.) Three characteristics of social prestige that I will account for and take into consideration are:

* That it is *distributed*: one has power/prestige if others think they have power/prestige; the power is in the perception.
* That it is *subjective*: a person does not have inherent social power; societies determine the distribution of social power. (This distribution can and likely will change and will never be “wrong.”)
* That it is *objectively knowable*: this is the crux/backbone of my research; that it is possible to determine popularity/power/prestige.

Due to the distributive and subjective nature of social power, it is important to recognize that social power is also recursive (as the more power you have, the more power you can bestow upon others, but you must be given power first to give power). This is why I will be using PageRank to determine social status and hierarchy, to account for its recursive nature.

**Domain & Approach**

My approach was from a technical/computational, data-science focused angle. I wanted to compile past social rank data and collect updated ranks and values to make a series of matrices that I could compute means, page-rank, and correlation data from. Because of the abundance of numerical data I collected, I was able to utilize a plethora of different data analytics techniques and find possible connections.

**Relevance to Theme**

I decided to compute page-rank to determine social rank/hierarchy, the crux of my paper, means to use a less “processed” metric for differences and correlation, and correlations to see possible connections, similarities, and differences between dimensions and metrics. Each of these gives me more insight into the way my small, distributed group functions and values its members.

Each of the questions I asked correlates to an important characteristic:

* How well do you know this person? (familiarity; allows me to find the most well-known person)
* How much do you like this person? (favor; allows me to find the most popular person)
* How often do you communicate with this person (via text, Snapchat, Facebook, in person, etc.)? (contact/communication; allows me to find the most connected person)
* How much do you trust this person? (trust; allows me to find the most trusted person)
* How likely are you to go to this person for advice or emotional support? (help/support; allows me to find the most helpful/possibly influential person)

Each dimension specifically points to an important person or important people in the community, but also may point to or correlate with one another. It is possible that every dimension is relevant to the others and may have some interaction with it. While this may be confusing and confounding, it may also provide some very interesting analysis.

**Method**

In order to collect the data and metrics needed to properly analyze the data and determine a social hierarchy in this group, I sent out an exhaustive (16 page) survey for each member of the group to complete. All 14 members of the group filled out the survey for the first time two years ago, between November 30, 2017 and December 6, 2017. 12 out of the 14 original participants filled out the survey for the second time between the dates of December 7, 2019 and December 9, 2019. The survey asked each participant to rate all 14 participants prompted by the following questions, and every question was required (their scores for themselves were ignored):

* How well do you know this person? (1 = not at all to 5 = very well)
* How much do you like this person? (1 = not at all to 5 = very much)
* How often do you communicate with this person (via text, Snapchat, Facebook, in person, etc.)? (1 = never to 5 = very often)
* How much do you trust this person? (1 = not at all to 5 = very much)
* How likely are you to go to this person for advice or emotional support? (1 = very unlikely to 5 = very likely)

The full survey text can be found in the Appendix (1.1).

**Limitations & Capabilities**

While it was truly fortunate that I had such relevant data lying around, ready for use and analysis, when I first created the survey, I was unaware that demographic data should be collected at the end of the survey (not at the beginning) to avoid priming subjects, so I asked about demographic data at the beginning. For the sake of consistency, I kept the survey and its order the same, including keeping the order of the people being evaluated in the survey the sane, but that may have introduced some unwanted error into the data. In addition, the some of the questions are not worded as clearly as I would like, and the scale is incredibly relative. (But having two sets of data does allow me to look at the difference and not just the raw values.) There are no open-ended questions, which might have allowed me to understand and analyze the data more holistically. In addition, I am a member of the group I surveyed, and these people are my friends. They knew I would be analyzing this data when taking the survey, and even though I asked them numerous times to be as honest as possible, they may have not wanted to be completely honest in scoring/ranking me or others in the group. These are some of the limitations of the collection method.

However, having two sets of data for the same friend group allows me to see and analyze the changes over time. The questions I had originally asked are very much relevant to the questions I am trying to answer and the problems I am trying to solve, as they are directly related to community, asking for help, popularity, familiarity, and trust—all important aspects of every community and social system. In addition, I had a willing set of participants (even though I have not talked to them recently, as you will see in the results!). Collecting data among five dimensions allows for so many different combinations and possible interactions/correlations. I see this as both a limitation and a great benefit. While I am not able to perform every possible computation and may be missing certain insights embedded deep within the data and this does complicate and lengthen the analysis I will be doing, it also provides me with a more robust data set, especially given the small group/sample size.

**Implementation**

Figure 1: The Technical Journey

Once a sufficient amount of data had been collected, I began the process of analyzing it. I collected a total of over 1,600 relevant, usable data points (not including demographic data or self-scores), and compiled all this data into 9 different Jupyter Notebooks, each focused on a different angle or of the data:

1. PageRank for 2017 Data\*
2. PageRank for 2019 Data\*
3. Differences Between PageRank Data\*
4. Means for 2017 Data\*
5. Means for 2019 Data\*
6. Differences Between Means\*
7. PageRank for Outward Data\*
8. Means for Outward Data\*
9. Cleaning & Compiling Data for Export to Excel

\* includes rankings, charts, gender differences, correlations, heatmaps

Doing analysis and creating visualizations in python were fast and simple, and they were incredibly helpful for uncovering and revealing the insights the data held. I used tables to display simple values and data, bar plots to show values and rank, and heatmaps to show possible correlations more clearly.

In Tableau, I was able to make more sophisticated visualizations, such as matrices of scatterplots with trendlines to quickly scan for interesting correlations.

All data, notebooks, and analysis can be found at <https://github.com/stephkananth/bubbles> for more details.

**Alternate Explanations & Confounds**

While doing the analysis, one the first criteria I filtered/grouped on was gender, as it sometimes shows significant differences and generally provides an interesting split. However, in this case, because of the small sample size, filtering on gender was not helpful, so I had to think of other lenses through which to see the data.

Also, due to the nature of the group (small and distributed), being the most liked or connected does not mean much and might not even be a good thing, as it may be an indication of a lack of depth in college community and post-high school friendships.

One thing that is important to remember in doing the analysis was that correlation does not equal causation, even if a 2017 dimension is highly correlated with a different 2019 dimension. There are many possible confounding variables that I did not take into great consideration including college/location/proximity, initial friendship level, and significant events that may have occurred. In addition, every member of this friend group voluntarily chose to be part of the group, so there are common experiences and interests members of the group have that are not taken into account by the survey. Others may have been forced into more “leadership” type roles in the group due to having a car or nice backyard.

**Results**

**Analysis**

**Significance**

**Appendix**

* 1. Survey

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated