Diversity and Popularity in Organizations and Communities

WALID NASRALLAH   walid@alum.mit.edu

PETER GLYNN   glynn@leland.stanford.edu

RAYMOND LEVITT   rel@ce.stanford.edu

Stanford University, Stanford, CA, 94305

Abstract.   This paper examines social groupings whose structure depends only on the distribution of ability to attract attention. When people's attention is a scarce resource, those individuals who are rated highest by a large number of other individuals will have to ration their attention, resulting in constraints on the social structure of the group. The incidence of popular individuals by that definition reflects the extent to which individuals agree on who their highest-rated colleague is. We propose three basic distributions or ways to generate the matrix of perceived ability so as to yield popularity profiles that can be parametrically adjusted to match observations. We demonstrate that each of these assumption sets leads to a slightly different correlation between the value of the assumption's parameter and the set of observable popularity patterns. Since popularity, in real life, often determines such things as power, centrality, over-utilization and perhaps reduced accessibility, having more realistic ways of representing it is important for modeling and understanding virtual organizations and communities.

Keywords: virtual organization, adhocracy, fully-connected networks, bounded rationality, coordination, communication, interaction, evaluation, centrality, popularity, organizational structure, social network genesis

1.   Introduction

Communication technology has already significantly changed the character of many work-places and organizations. Increases in bandwidth, automation and end-user accessibility are making it almost as easy to interact with someone halfway across the planet as with someone across the hall. Multinational corporations have proprietary phone networks. Universities maintain World Wide Web listings of their faculty's research interests and publications for the benefit of the global academic community. Communities that form around Internet chat rooms and Usenet groups demonstrate that the potential power of new communication technology can already supplant face-to-face conversation for many purposes, including genesis, maintenance, and evolution of social groupings. Large economic benefits can be realized by firms and other organizations that use these communication technologies to their full potential. However, there are also costs that are not fully understood. Perhaps chief among these is the increase in individual workloads of managers and workers concomitant with decreases in the limits on possible interactions with colleagues, clients, suppliers, competitors, and so forth. As the pace of these changes increases, future work in organizational modeling and re-engineering will need a new basic metaphor for understanding how to balance these benefits and these costs.

   A metaphor that comes to mind is that of a "mixer" cocktail party where all the guests are equally unfamiliar with one another, and they can see and instantly decide on how interested they might be in approaching certain individuals. The instant evaluation mirrors the ability to look up information on-line about a potential business or research partner.

Presence in one large ballroom mirrors the instant accessibility afforded by email today or by holographic teleconferencing in the future. What would be the dynamics and patterns of social interactions in such a situation? Would there be a big clique around the most attractive person, or will the circles around the second- and third- most attractive be bigger because too few people believe they stand a chance of getting through to their favorite? Where do the social dynamics that emerge from such a model agree with and where do they differ from existing sociometric models that are built on assumptions of relative scarcity of, or constraint on, social opportunities? And finally, are there any recorded observations available that can be seen to match these predicted patterns today?

Note that these questions can just as easily arise in a completely non-frivolous manner under certain professional settings. For example, we can think of a complex construction site with several dozen contractors, designers, sub-contractors and sub-designers all head-quartered in a large cluster of trailers on the site. Invariably, each of these contractors will face daily issues that require communication with some other party: problems of access, coordination, missing information, underground surprises, late changes, code revisions, and so on. Most of the parties will have a binding contract with either the project owner or with some intermediary which has a such a direct relationship to the owner. Since it is seldom politic to engage the owner in each tiny altercation, formal organizational relationships quickly become negligible for most communications. Similarly, the long duration and limited geographical extent of the project, coupled with a need continuous low-level coordination between all project participants, quickly equalize accessibility and neutralize the effects of past cooperation on projects. Viewing these interactions at a high level of abstraction, we can say that for any randomly chosen day, the chances that each firm will benefit from an exchange with some other firm can be grossly estimated from some knowledge of the division of labor, the site layout, and the phase of project on that day. What we cannot directly deduce is how rapidly each party will be able to respond to those requests. Who winds up being the most reliable source of resolving a conflict? Not necessarily the one with the most power to resolve the problem, but rather the one who has enough power to make a change *and* who is also free enough or interested enough to take action. But the more action one chooses to take, the less free time one has to respond to new requests. Just like the cocktail party, the patterns of interaction that emerge will be complicated by the presence of feedback loops that in some cases may overpower the initial distribution of popularity and shift things over to a different pattern of interaction.

We begin our investigation of such low-constraint social networks by building upon an existing mathematical model developed by Huberman and Hogg (1995) to model communities of practice. First we describe the common abstraction that links the different domains we have discussed, and makes it possible to make predictions about all of them from one simple mathematical model. The notion is that similar interaction patterns tend to emerge when everyone chooses to exercise the ability to talk to some people (and ignore others) based on the perceived intrinsic value of the communication. In other words, we are not using the network to depict a status quo, or to represent capacity for transmission of some type of information or other benefit from one point to another(Scott, 1991). Neither is the goal to increase network connectivity through establishing new links based on common acquaintances as in (Carley, 1990, e.g.). Our generalization of the Huberman & Hogg (1995) metaphor uses the same graph-theoretical language as many social network analysis

paradigms. This commonality of language makes it easy to productively mix the various models. For example, the shape of the optimum network obtained by a popularity/ability model might be seen as a starting point for network evolution (Banks and Carley, 1996, e.g.) under increased interaction, or it may be seen as a steady state end point for the evolution of a network which had evolved under high-constraint circumstances. More complex patterns might involve simultaneous application of forces from different published models that influence future affinity. Finally, linear algebraic methods such as singular value decomposition might be used, as suggested by (Freeman, 1997), to translate the interaction matrices into hierarchies of dominance, creating a bridge to the less mathematical bodies of research in organizational behavior.

That said, we list below the basic distinguishing characteristics of this distinct modeling metaphor.

- The interaction *per se* generates something of value to the participants. Unlike other social network models, the value is intrinsic to the two parties interacting, and is not related to their formal role or their degree of connection to third parties.

- There are no barriers of any kind between people - once both parties decide to pursue an interaction, no additional requirements need to be met.

- There exists a way of measuring the value of each possible interaction. This value might be fixed or changing, and it might be partially unknown to the participants as they make their decisions, but we assume there is some way of estimating it at any point in time.

- Interaction value may or may not be correlated to the frequency of past or future interaction. In cases where it is, the social network concept of "tie strength" would be a close analogue.

- A past interaction between two individuals may or may not affect the value of future interaction between other individuals. Such an effect may be completely arbitrary, as in (Huberman and Hogg, 1995, section 4), or it may be in part determined by an underlying logic as in the Constructural theory of (Carley, 1991).

These assumptions are complementary. For example, in a network that models knowledge transmission, being with a good teacher helps one become a good teacher, but in a network that models coordination on a construction project or banter at a party, popularity is a function of inherent, non-transferable traits. Similarly, if there are structural impediments that prevent you from going to your favorite friend, you might value the company of someone who has been with that friend recently, but if you can more easily go to the friend than to the third party, then the third party will not gain any value in your eyes from their past interaction with the friend. There is some evidence (Brass and Burkhardt, 1992) that the degree of connectivity, calculated purely by counting direct links, is at least as useful as any more complex measure in predicting organizational effectiveness and individual power in organization. Models that include only direct connectivity therefore have a role to play in relating relevant environmental factors to observed interaction patterns. Such models may or may not make the additional strong assumption of independence between past interactions and the attribute values that determine future interactions. We will, however, focus on the simplest possible case in this initial investigation.

2. Background

In section 3 of the Huberman and Hogg (1995) paper, there is a chart entitled "Clustering as a Function of Diversity and Size of a Community of Practice". This chart is an archetype for a class of general predictions that can be made about social networks that arise in real life by virtue of being "optimum" in some way under certain contingency factors. Behind that chart is a social network model that, unlike many others (Carley, 1990), evolves towards a degree of connectivity far below full connectivity. The model is based purely on perceived attributes of individuals, which according to (DiMaggio, 1992) is only one of three factors that determine the shape of a social network. The other two are formal organizational role and pre-existing connections to others. Focusing on a single mechanism for evolution of a network allows us to better understand that one factor, as long as one is aware of and can compensate for the effects of the other factors. We believe that the onset of virtual organizations is one way in which the effects of organizational roles and of pre-existing network relations will be reduced in an observable real-life setting. We also agree with (DiMaggio, 1992) that situations where these factors are also present can still be best analyzed based, first, on a detailed representation of each of the three factors in isolation, and only afterwards on how the three factors interact when combined.

We therefore now focus on determining which properties or components of the Huberman and Hogg (1995) model need to be duplicated to model different organizational settings with purely attribute-based interaction, and what other ways exist of putting together models that have that property. This will allow us to vary those modeling assumptions that are still free, and thereby develop a slew of specialized, simple models that can be tested for predictive power under various circumstances. Such models might, for example, be used to:

- Evaluate a specific organization for good fit with measured parameters of the environment.
  e.g. do retail bank employees need more access or less access to their peers in the organization than investment bankers?

- Decide on types of changes that need to be made in a presently successful organization that faces predictable changes in its environment.
  e.g. does the institution of medical care protocols by health management organizations mean that doctors will need to talk to one another less frequently?

- Predict the effects of increased accuracy of information about other members on the number or pattern of interactions that take place.
  e.g. how much productivity is lost when people of different ethnic backgrounds have unfounded worries about one another's competence? What if only one side is misinformed and the other has a true picture of the first's abilities?

- Give a very specific and simple explanation of why a certain type of organization is prospering in a certain industry.
  e.g. Suppose we observe that centralized Health Maintenance Organizations (HMO's) are stamping out decentralized HMO's of equal size. We might explain that with a simple two-parameter model, where the degree of specialization of doctors and the degree of differentiation in illnesses are the two model inputs. If technology or demographics

$$
\begin{bmatrix}
s & 1 & 1.1 & 1.21 & 1.331 & \boxed{1.4641} \\
1 & s & \boxed{1.4641} & 1.331 & 1.21 & 1.1 \\
1.331 & 1.21 & s & \boxed{1.4641} & 1 & 1.1 \\
1.331 & 1.1 & 1 & s & \boxed{1.4641} & 1.21 \\
1.21 & 1 & 1.1 & 1.331 & s & \boxed{1.4641} \\
1.331 & 1.21 & 1.1 & \boxed{1.4641} & 1 & s
\end{bmatrix}
\quad
\begin{array}{l}
\text{Ratings produced by person 1} \\
\text{Ratings prodiced by person 2} \\
\vdots \\
\vdots \\
\vdots \\
\vdots
\end{array}
$$

Figure 1. A Sample Cross-utility Matrix (highlighting the favorites)

> *have shifted the parameters in a certain direction, and the model predicts a change in the optimum degree of centralization, then we have a testable explanation.*

- Determine values of hard-to-measure variables of the social environment based on the preponderance of firms with certain structure in that environment.

The predictive power of these models will be somewhat diminished in the presence of various physical and institutional barriers to interaction. Attributes of networks and of individuals within a network often depend on precisely those impediments. This much written-about premise is neatly expressed by the idea of "Structural Holes" (Burt, 1992) which, briefly stated, attributes organizational power to an individual's ability to act as an intermediary between others who do not have a direct connection. However, we argue that, whereas it is easy to imagine a world where the barriers are all removed and interactions depend solely upon individual preferences or pairwise cross-utilities, it is hard to imagine a world in which there are only barriers and no preferences. Clearly, both sorts of models are needed to explain complex social phenomena in the real world, where behavior is influenced by structural holes and by patterns of preference, as well as by other factors which may be random or just not well understood.

3. Premise

In the original (Huberman and Hogg, 1995) model, each member of the group had a separate rating for all the other members, with a favorite, a second-favorite, and so forth. The model thus embodies several simplifying assumptions:

1. Only one person occupies each rank (i.e. no one has two or more equal favorites)

2. A *real-valued* measure underlies this rating (i.e. it makes sense to say "X is twice as valuable as "Y".)

3. The ratio of the nth-best to the $(n-1)^{th}$-best is constant for all n (i.e. for all people being rated).

4. The same ratio of nth-best to $(n-1)^{th}$-best is used by all people who are doing the rating.

$$
\begin{bmatrix}
s & 0 & 0 & 0 & 0 & 1 \\
0 & s & 1 & 0 & 0 & 0 \\
0 & 0 & s & 1 & 0 & 0 \\
0 & 0 & 0 & s & 1 & 0 \\
0 & 0 & 0 & 0 & s & 1 \\
0 & 0 & 0 & 1 & 0 & s
\end{bmatrix}
$$

Figure 2. Matrix of Favorites Highlighting the Most Popular Persons

5.   Each person's ratings are completely independent of each other person's.

In effect, the results obtained by Huberman and Hogg relating network clustering to network size and variance come out of assuming that the matrix of perceived cross-utility ratings is populated from a geometric series, e.g. 1, 1.1, 1.21, 1.331, 1.4641, ..., 1.1^n, and each row is a randomly permuted sequence of the same series. A matrix of crosswise ratings might thus resemble Figure 1. A consequence of having such a distribution, which we prove in Appendix B.1, is that the most popular person in the group is expected to be rated "best" by only a very small proportion of the population. For clarity we stress the following distinction: **X's Favorite**: the person receiving the top rating by person X. Everyone has only one favorite in the original Huberman/Hogg model. **Y's Popularity**: the number of people overall who have chosen Y as their favorite. For a population of "n", the modal assignment of preferences has a number of people choosing the most popular as their personal favorite is given by:

$$
\left\lfloor \frac{\ln(n)}{\ln(\ln(n))} \right\rfloor \tag{1}
$$

For a population of 5 billion, this number is 7! This might be a plausible number when we are dealing with the world of scientists, where ideas cross-fertilize across narrow fields of knowledge, and everyone is a "leader" in one of these fields. But to model most organizations, communities, or societies, we need to make slightly different assumptions to account for the high incidence of popular or acclaimed individuals who are favored by a majority of the population.

A model that admits variations in popularity is especially necessary when the model output depends on this particular measure. For example, we can imagine many situations where being popular leads to being busy: socialites at a party, cardiologists at a hospital, artists at an advertising agency, diesel mechanics at a construction company, and so on. In this paper, we are not concerned with the exact mechanism of how popularity translates to over-utilization or to unavailability. Instead, the focus is on how to align the model's measure of popularity with observations, in preparation for building more advanced models. We simplify by looking at the top-rated individual versus everyone else, despite awareness that being second-favorite to many might in some situations make one more busy than being favorite to a few. The results about popularity are interesting in their own right, and are

also directly applicable when modeling organizations with a large distinction between the favorite and the second-favorite. In models with small distinction between the first and second, the results below are merely indicative of what would be obtained with additional mathematical treatment which we hope to undertake in future research.

4. Alternative Model Assumptions

We would like to allow the popularity patterns in the model to be fitted to observations of a particular population or organization[1]. There are two ways to do so without losing the other simplifying assumptions of the model. One is to give each person two ratings - a universal one which makes him or her more skilled/attractive/valuable to all others, and a residual one which is specific to each of the other members. The other is to use a small number of independent rank lists, from among which every individual must pick[2] one or more.

The former, which assigns an ad-hoc homogeneity measure based on observed popularity, is the simplest, but the latter, with its multiple criteria model, is more general. The two can be combined with one another, as well as with a third modification where ratings are defined over percentiles instead of on absolute numbers.

The following three sections examine the mathematical effects of each of these three alternative assumption sets on the relationship between popularity and the relevant model parameter. We continue to factor out the effect of variance in the absolute ratings by only considering the distinction between "favorite" and "other", where "other" encompasses everyone from the second-favorite to the least favorite, and where being "most popular" means merely being a favorite of the greatest number of people.

4.1. Degree of Homogeneity

If the whole population is fully homogeneous in their tastes, then everyone will clearly favor the most popular, and the most popular will be preferred partner of everyone else in the group. By contrast, the result we obtained before rests on an assumption of full *heterogeneity*. The suggestion in this section is to take a linear combination of the two cases, and thus have a model where the proportion of people favoring the most popular can be arbitrarily set to match observed data by varying what we call the degree of homogeneity in tastes.

We illustrate the relationship between percent homogeneity and proportion favoring the most popular using the following example:

Suppose that everyone in the country has an opinion about who the most intelligent among them is, and that 70% of all people believe that the Guinness Book of Records is the single authority on who that person is. Marilyn vos Savant would thus be selected by all the 70% who follow the Guinness book. The other 30% will choose someone arbitrarily, and some of them will by chance also choose Marilyn. We may wish to maintain the sparsity of the model by continuing to simplistically assume that the 30% make their choice in a random manner, hoping that the effect of the assumption on the results will be smaller than the range of observational error. If we do so, the proportion of the 30% for whom Marilyn is also the most intelligent will be obtainable by a calculation similar to the one performed in Appendix A. Since we assume independence not only within the 30% but also

between that segment and the homogeneous 70%, the person most often selected as most intelligent within that 30% segment will almost certainly not be Marilyn. We expect that person to be selected $\left\lfloor \frac{\ln}{\ln(\ln)} \right\rfloor (30\% \times n)$ - if n is 250 million, he or she would be selected by $\lfloor \ln(75 \quad million) \div \ln(\ln(75 \quad million)) \rfloor = 4$. Since this person may or may not be Marilyn vos Savant, 4 is an upper bound on the number of people in the heterogeneous group who select Marilyn. This number is clearly negligible compared to the 175 million who chose Marilyn based on the Guinness book. However, in a group the size of a small organization, the ratio becomes more even and it becomes important to make a more exact assessment of how often the heterogeneous component chooses the same person as the homogeneous group.

Assuming that the favorite of the homogeneous segment is chosen independently of the choices of the heterogeneous segment, we need to look at the number of people in a fully heterogeneous population who rate a *randomly selected* person as their favorite. This number is most likely to be exactly one. As the population gets big, the number of people getting exactly one vote will approach e^{-1} from above, and the number who get no votes will approach e^{-1} from below. Only $1 - 2\,e^{-1}$ will get more than one vote. In short, the adjustment between degree of homogeneity and proportion favoring the most popular will be 1 or less.

Since the effect of the non-homogeneous portion of the ratings is so small compared to the homogeneous portion, we do not need to worry about its exact distribution. Hence, for a particular instance of the model, we only need the homogeneous ratings given to each individual, and the proportion of individuals who subscribe to the homogeneous rating system. This would be true even if the number of subscribers to the homogeneous system is a small proportion of the total population, unless that proportion became comparable to $\left\lfloor \frac{\ln(n)}{\ln(ln(n))} \right\rfloor$. Finally, the model would also undergo no change if we interpret the degree of homogeneity measure as a weight that each individual gives to the homogeneous rating when performing an evaluation of another.

We note here that models built upon the single "degree of homogeneity" measure can help future investigations of such issues as variation in homogeneity over time, e.g. whether and how media homogenize our ideas of beauty, schools homogenize our ideas of intelligence, and global commerce homogenizes our ideas of morality.

However, things do become different if we assume that there are multiple competing rating systems that claim several adherents; this is what we examine in the next section.

4.2.   Limited Evaluation Criteria

It seems too arbitrary to assume that there should be one global rating system and several independent individuals. What if there are several rating systems, or perhaps several rating criteria, each of which was followed by a small number of individuals? Instead of being concerned with how many people follow the global system, we now need to ask "How many systems are there to choose from?" This number has a determinable mapping to the proportion choosing the most popular individuals, which we derive in Appendix B.

But first let us look at an example that illustrates the generality of this class of assumptions. Suppose that a population of 10 people has a choice between three evaluation criteria. We

can represent the situation by a 10-row, 3-column matrix that lists the relative relevance to popularity assigned by each member to each criterion, and a 3-row, 10-column matrix that shows how each member scores based on each of the three criteria. Even if we assume that the criteria are independent (in probabilistic parlance) or orthogonal (in linear algebraic terms), there will still be significant information redundancies in the 10-by-10 product of these matrices, which plays the role of the regular preference matrix (just like the one in Fig. 1). These redundancies reveal themselves in the vanishing determinant of the 10-by-10 matrix. In fact, the determinant of any sub-matrix of that matrix larger than 3-by-3 will also be zero. The information we need to judge the patterns of popularity inherent in the 10-by-10 matrix is more concisely represented in the column sums of the 10-by-3 relative relevance (or weight) matrix, and the 3-by-10 matrix of attribute scores. If we have a situation with a single global rating system that is followed by 8 out of the 10, the matrices might look something like this:

$$
\begin{array}{ccc}
Weight & \times & Rating
\end{array}
$$

$$
\begin{bmatrix}
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\times
\begin{bmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & \boxed{9} & 0 \\
5 & 3 & 2 & \boxed{9} & 0 & 8 & 1 & 4 & 7 & 6 \\
7 & 6 & 1 & 8 & 2 & 3 & 0 & 4 & 5 & \boxed{9}
\end{bmatrix}
\tag{2}
$$

We see that the first attribute receives a weight of one from eight out of the ten members, so it has a cumulative weight of 0.8, which is all we need to know. If we are looking at the most popular individual, this will clearly be the one with the highest score for that attribute, which in the above example will be the 9th individual. The same result would be obtained if we changed the weight matrix as follows:

$$
\begin{array}{ccc}
Weight & \times & Rating
\end{array}
$$

$$
\begin{bmatrix}
.8 & .1 & .1 \\
.7 & .1 & .2 \\
.6 & .2 & .2 \\
.9 & 0 & .1 \\
.7 & .2 & .1 \\
.8 & .2 & 0 \\
.8 & .2 & 0 \\
.9 & 0 & .1 \\
.8 & 0 & .2 \\
1 & 0 & 0
\end{bmatrix}
\times
\begin{bmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & \boxed{9} & 0 \\
5 & 3 & 2 & \boxed{9} & 0 & 8 & 1 & 4 & 7 & 6 \\
7 & 6 & 1 & 8 & 2 & 3 & 0 & 4 & 5 & \boxed{9}
\end{bmatrix}
\tag{3}
$$

Now, we no longer require each individual to religiously adhere to one and only one rating system. This reflects the observation that people who need help from others often need that help in different domains. For example, some may need a large proportion of technical help and a small proportion of communication help, or perhaps different mixes of certain resources commanded by others. Note that in this example, the column sums of the weight matrix are still the same, and the attribute value matrix is unchanged. As before, the popularity of each person is based only in the rating he or she receives based on the first attribute (i.e. the first row of the rating matrix.) But again, this is not always the way things are in the world we want to model with these methods. We might, for example, encounter cases where the different attributes have equal weights. This might arise because of a weight matrix such as the one below:

$$
\begin{matrix} Weight & \times & Rating \end{matrix}
$$

$$
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \times \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & \boxed{9} & 0 \\ 5 & 3 & 2 & \boxed{9} & 0 & 8 & 1 & 4 & 7 & 6 \\ 7 & 6 & 1 & 8 & 2 & 3 & 0 & 4 & 5 & \boxed{9} \end{bmatrix} \tag{4}
$$

Again, we have kept the same rating matrix, but now, the focus shifts to the properties of that matrix as a whole, rather than a single row. Instead of looking at the Guinness book versus individual opinions (in the words of the example,) we are now looking at equally visible and equally global attributes. In a business setting, we might be looking at different skill types or degrees of experience in doing some sort of work: negotiation, number crunching, and writing, for example. In a social setting, these would be nominally independent measures of social attractiveness – perhaps wealth, beauty and perspicuity. Clearly, since different people value these attributes to different extents, we cannot rely on just one attribute when predicting how popular people will be. These dimensions are analogous to Kathleen Carley's "facts" in the Constructural model (Carley, 1991, Kaufer and Carley, 1993). Unlike the Constructural model, however, these attributes cannot be acquired by contact with someone who has them. We can use this approach to determine how many such distinct rating scales exist in a population, based on observations of clustering patterns and estimates of the ratio between best and second best (which we still assume to be constant).

As before, the question of who will be most highly rated is open to investigation and depends on the ratio and distribution of the ratings for each attribute. We take a first stab at the problem by again simplifying things down to the case where we only care about the highest rating. As before, the matrix becomes one of ones and zeros, as follows:

$$Favorites \quad = \quad \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

In particular, we need to single out people who are rated highest in more than one attribute, since these are the ones who are likely to be the most popular when the attributes are equally weighted by the population. More generally, we need to do so even for unequal weights as long as no single attribute is sufficient to overwhelm a combination of two or more others. In this matrix formulation, we care about the possibility of there being more than one "1" in any column, and if so, what is the number of "1"s in the column with the most. We use an asymptotic argument (as $n \to \infty$) to show that the relevant parameter is the relationship between the number of attributes and the number of individuals in the population. Briefly, the results of the mathematical investigation reveal that the number of rating systems may fall in one of three ranges:

1. Constant or proportional to less than the square root of the number of people in the population.

2. Proportional to the population count.

3. Proportional to some intermediate power of the population count.

The first case turns out to be extremely unlikely to yield matrices with multiple "1"s per column. This means that if the number of attributes is much smaller than the square root of the number of people, there is a very slim chance that anyone will score the top rating on more than one attribute. The second case is identical to the original proposition of completely independent individual ratings; the highest number of top-ratings anyone can reasonably be expected to get is about $\left\lfloor \frac{\ln(n)}{\ln(\ln(n))} \right\rfloor$. This is multiplied by the number of people per rating system to yield the number of people who gave the most popular person their vote for favorite. The last case yields an intermediate number which is related to the power relationship between n, the number of people in the population, and m, the number of attributes. The expression is $M_n = \left\lfloor \frac{1}{1-p} \right\rfloor$ where p is defined by the relationship $m = n^p$, and M_n is the number of *rating systems* which give a top rating to the most popular individual.

For most approximation purposes, treating a finite value of m as proportional to the finite value of n is all that we need. For example, if we have 9 equally subscribed rating systems in an organization of 30 people, then we can expect the most popular person to have

$$\frac{30}{9} \times \left\lfloor \frac{\ln(30)}{\ln(\ln(30))} \right\rfloor = 6\tfrac{2}{3} \cong 7$$

It is possible to extrapolate a value for the number of equally weighted rating systems given a proportion of the population voting for the most popular, consistent with the assumption that the number or rating systems m is proportional to the total population n. With more data points on popularity of various individuals, one can vary the weights of the different rating

systems to match the data points without abandoning the hypothesis of proportionality between m and n. But for completeness, let us examine the same situation we just did under the assumption that the 9 rating systems are actually derived from $n^{0.646}$. Then the asymptotic approximation for the number of people choosing the most popular as their favorite is:

$$\frac{30}{9} \times \left\lfloor \frac{1}{1 - 0.646} \right\rfloor = 6\tfrac{2}{3} \cong 7$$

The two are the same in this range of values, but they diverge for different values, giving us more flexibility in modeling.

4.3.   Number of Equal Winners

All of the results discussed thus far depend on the property that there is always a single top-rated individual for each rating agency, system or attribute. Things change when we posit that the ratings are based on percentiles of the population, so that, for example, all individuals in the top one-tenth of the population get the same top rating. The operative difference is that, under this policy, doubling in population size leads to a similar doubling of people assigned a "number one" grade by some rating system.

We start off by looking at the mean expected number of "grade one" ratings that any individual is expected to get. If the number of "grade one" ratings that each rating system gives is fixed at d_n, and there are n individuals being rated and m_n rating systems, then clearly we would expect the number of "grade one" ratings received per individual to be $(m_n \times d_n) \div n$. The Lattice Distribution Central Limit Theorem (Feller, 1968, 2:XV.5, pp. 517-518)[3] tells us that this average is also the most likely, because the number of ratings received by an individual will tend towards a Normal distribution. More importantly, the theorem gives us the appropriate variance for that Normal distribution, which allows us to easily determine the number of "grade one" ratings received by both the most popular and the least popular in the population. It turns out that the range is very narrow when the number of "grade one" ratings is proportional to the number of people (i.e. if the rating is a percentile-based rating). Specifically,

$$C_n(k) \approx \frac{m_n \, d_n}{n} + \sqrt{\frac{m_n \, d_n}{n}} \, N(0,1) \tag{5}$$

where $C_n(k)$ is the sum of the kth column in the ratings matrix (i.e. the number of "grade one" ratings received by the kth individual,) and $N(0,1)$ is a Normal distribution with 0 mean and 1 standard deviation. If we define M_n as the maximum of the series $C_n(k)$, then

$$M_n \approx \frac{m_n \, d_n}{n} + \sqrt{\frac{m_n \, d_n \, 2\ln(n)}{n}} \tag{6}$$

and by symmetry of the normal distribution, the minimum column sum will be given by a similar expression with a minus sign in front of the square root.

4.3.1. Numerical Example Suppose that we have a 100-person consulting firm where every person rates the others' usefulness independently of all other, using a scale with 10 possible ratings. If we extrapolate the asymptotic result back to the $n = 100$ point and assume it has some veracity, we would expect the most highly acclaimed individual will be given the top rating by

$$10 + \sqrt{\frac{10 \times 2 \times \ln(100)}{100}} = 10.96 \cong 11$$

people. If, on the other hand, there are only five independent rating systems which are equally subscribed to by all individuals, then the number becomes:

$$20 \times \left(\frac{50}{100} + \sqrt{\ln(100)} \right) = 20 \times \left(\frac{1}{2} + 2.15 \right) \cong 53$$

whereas the person with the least number of "grade one" votes will have none (because the equation yields a negative number, but the distribution is truncated at zero.)

4.4. Result Summary

The following table relates the various model parameters discussed above to observable organizational popularity/centrality information. The first four columns list the parameters and the last four list the most likely popularity rating for an organization with these parameters. The table is not comprehensive, as we have left out the more complex expressions that result from esoteric combinations of parameter ranges.

Note that i, j and k in column 6 refer to the different criteria that give a top rating to the same person. That person becomes the most popular only if the sum of the weights of those criteria exceeds the weight of the most heavily subscribed criterion. Otherwise, the weighting of different criteria might be so unequal that a singly heavily subscribed criterion decided who is most popular.

5. Implications for Diversity

Huberman and Hogg (1995) used the word "diversity" to refer to the ratio between the real number attached to the top rating and the real number attached to the bottom rating, assuming that all intermediate ratings are evenly spaced in some way. In the course of the work above, we have introduced two new concepts which can also be denoted by the word "diversity". We factored out the effect of the original diversity measure by focusing solely on the distinction between the top-rated individual and everyone else. The two new measures of diversity that emerged were, first, the number of independent ways in which people are rated, and, second, the number of distinct ratings that people can get within a rating system. We can next proceed to examine the effects of varying these different measures of diversity on the emergent network of actual interactions. These can be charted from pure mathematical calculations, as discussed in section 6. What is more significant is that we have tied the two new measures of diversity to another statistic which may in practice prove easier to collect: the collective rating value of most popular and least popular

No. of Equal Firsts in Criterion	No. of Indep. Criteria (m)	Group Size ÷ Criteria $\left(\frac{n}{m}\right)$	Weights of Criteria	Popularity of Most Popular	Popularity of Second	Modal Pop. (Most Likely)	Pop. of Least Popular
1	2 – 8	1	N/A	2	1	1	0
		r	equal	$2 \times r$	r	r	0
			descending: $z_1,\dots,z_8$	$\max(z_1,(z_i+z_j))$	2nd highest of $(z_1,z_2,(z_i+z_j))$	r	0
	9 – 5503	1	N/A	3	2	1	0
		$r \ll \sqrt{n}$	equal	$3 \times r$	$2 \times r$	r	0
			descending: $z_1,\dots,z_m$	$\max(z_1,(z_i+z_j+z_k))$	2nd highest of $(z_1,z_2,(z_i+z_j+z_k))$	r	0
	> 5503	1	N/A	$\frac{\ln(n)}{\ln(\ln(n))}$	$\frac{\ln(n)}{\ln(\ln(n))}-1$	1	0
		$r \ll \sqrt{n}$	equal	$r \times \frac{\ln(n)}{\ln(\ln(n))}$	$r \times \left(\frac{\ln(n)}{\ln(\ln(n))}-1\right)$	r	0
	> 10	$r \approx \sqrt{n}$	equal	$r \times 1$	$r \times (l-1)$	$< r$	0
	> 10	$r \approx n$	equal	r	≈ 0	0	0
$d < n$	> 10	$r \mathbin{\text{¿}} 1$	equal	$(dr)+\sqrt{dr\,2\ln(n)}$	Gaussian Dist.	dr	$(dr)-\sqrt{dr\,2\ln(n)}$

Table 1. Summary Table

individuals. It should also be theoretically possible to tie in the first definition, the rating slope curve, to similar statistics if we re-derive the popularity results from rating matrices that have more than ones and zeros.

6. Future Work

This paper has been concerned with only half of the model - the rating value matrix which gives the value of an interaction subject to the successful completion of the interaction. With the understanding gained thus far of the different possible ways of structuring this value matrix, we plan to tackle some variations of the original model which operate on the principle of maximizing

$$E = \sum_i \sum_j H_{ij} \times P_{ij} \times S_{ij}$$

Where

- H_{ij} is the value matrix which is the model's input,

- P_{ij} is the matrix of how frequently i chooses to go to j, and

- S_{ij} is the probability of success that i faces when going to j. In general, S may be a function of the whole P and H matrices.

We plan to introduce a new S_{ij} function which incorporates bounded ability to respond to interaction requests, and we will examine the effects of varying the different parameters

discussed above on the shape of the P matrix when the P matrix is arranged to maximize the global efficiency. This process will yield several predictions about organizations which we will test against previously made observations. This will allow us to determine the applicability of this simple model under certain circumstances. As we discussed before, we expect to be able to account for some but not all patterns of behavior in organizations where the barriers to communication are removed or reduced.

Appendix A

Popularity by Counting

We assume that each person in a finite group has a definite single "favorite" in the group. We use the word "popularity" to denote the number of "favorite" ratings the a person has from others in the group. The most popular person in the whole group is thus the one who is chosen as the favorite my the greatest number of people. Let us start out with the simplest possible case to illustrate the principle. If there are two people who interact even occasionally, each will be the most popular partner of the other, because there is no one else. If there are three people and each chooses a favorite among the others at random, then there is a one-in-four chance that person A is selected by both B and C, and a one-in-two chance person A is selected by one but not the other, and a one-in-four chance of A being the favorite of neither. The process is the same for the popularity rating of B and C. If all are independent, then the combinations are as follows:

Table A.1. Table of probabilities of different combinations of favorites in a 3-person group

Number of ways that three people can pick one of two	$= 2^3 = 8$
P [Two or more are favorites of both others]	$= 0$
P [Two or more are favorites of none]	$= 0$
P [One favored by both others, who in turn favors just one]	$= \frac{6}{8} = 0.75$
P [Each person favored by exactly one]	$= \frac{2}{8} = 0.25$

Clearly, in this situation the most likely outcome is that the most popular person will get that distinction merely by having both others nominate him or her as a favorite. But it turns out that being chosen by just two people as a favorite is enough to make you the most popular in a group of up to 8. Even though it is possible to be chosen by 3 or more, the odds of that happening to anyone in the group will be low until the group size is 9 or larger. The actual numbers are tabulated in Table A.2. Note that we slightly simplify the mathematics by pretending that each person can choose among n instead of $n-1$ potential favorites.

The calculation gets tedious for larger numbers, since the complexity of working out the combinatorics of each possible outcome grows exponentially with the number of people in the system. The process for carrying out the calculation is given below[4] :

Define $S(k, n)$ as the number of different ways that n individuals can each select a single favorite out of the remaining $n-1$ such that the most frequently chosen (i.e. most popular) person in the whole group is chosen by exactly k people. Note that the number of people

Table A.2. Possible Assignments of Favorites in a Population of 4 – 8 People

No. of people choosing the most popular as their personal favorite	Possible Combinations where the Group Size is				
	4	5	6	7	8
1	24	120	720	5,040	40,320
2	180	2,100	28,800	458,640	8,361,360
3	48	800	14,700	301,350	6,867,840
4	4	100	2,250	52,920	1,342,600
5		5	180	5,292	153,664
6			6	294	10,976
7				7	448
8					8
Total	256	3,125	46,656	823,543	16,777,216
Combinations	$= 4^4$	$= 5^5$	$= 6^6$	$= 7^7$	$= 8^8$

attaining the "most popular" rating can be more than one.

$$S(k,n) = \sum_{i_k=1}^{q(k,n)} f(k,n) \sum_{i_{(k-1)}=0}^{q(k-1,n)} f(k-1,n) \ldots \sum_{i_2=0}^{q(2,n)} f(2,n) \frac{(n - \sum_{m=2}^{k} i_m)!}{(\sum_{m=2}^{k} m\,i_m)!} \quad \text{(A.1)}$$

Where

$$f(j,n) = \binom{c(j+1,n)}{i_j} \prod_{m=0}^{i_j-1} \binom{(b(j+1)-m\,j)}{j}$$

$$q(j,n) = \left\lfloor \frac{b(j+1)}{j} \right\rfloor$$

$$c(j,n) = n - \sum_{m=j}^{k} i_m$$

$$b(j,n) = n - \sum_{m=j}^{k} m\,i_m$$

The number we want is the maximum of all $S(k,n)$ for all the k's from 1 to n. Since the number of nested summations is linear with the number of people in the group, the calculation grows exponentially with group size. For this reason, we resort to an asymptotic approximation, as described below.

Appendix B

Asymptotic Approximation of the Growth Rate of Popularity

We know that each person in the population has exactly one favorite, and this favorite could be anyone in the population. We can better visualize this as an $n \times n$ square matrix with one

row for each person, who has one "vote" to give, and a column for each person, in which an arbitrary number of votes (between 0 and n) may fall. There is only one vote in each row, and that vote is equally likely to be in any of the n positions in that row. Furthermore, we recognize that we might in future be faced with a situation where the population doing the rating might be different from the population being rated, so we remove the requirement for a square matrix and instead formally state the problem as follows:

1. Matrix $A = (A_{jk} : 1 \leqslant j \leqslant m, 1 \leqslant k \leqslant n)$

2. The rows $(A_{jk} : 1 \leqslant k \leqslant n)$ are independent, identically distributed (i.i.d) random n-vectors.

3. For all rows (i.e. for $1 \leqslant j \leqslant m$), $(A_{jk} : 1 \leqslant k \leqslant n)$ is multinomial $\left(1; \frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$

Definitions:

$$C_k \quad = \sum_{j=1}^{m} A_{jk} \qquad \text{i.e. sum of the } k^{th} \text{ column.}$$

$$M_n = \max_{1 \leqslant k \leqslant n} (C_k) \qquad \text{i.e. maximum column sum.}$$

The vector $(C_1, C_2, \dots, C_n)$ will be a multinomial $\left(m; \frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$, because it is the sum of m i.i.d. multinomial $\left(1; \frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$ random vectors. It follows (see for example (Feller, 1968), 1:VI.9) that :

For $1 \leqslant k \leqslant n$

$$E[C_k] = \frac{m}{n} \tag{B.1}$$

$$Var(C_k) = m(\frac{1}{n})(1 - \frac{1}{n}) \tag{B.2}$$

$$P[C_k = j] = \binom{m}{j} \left(\frac{1}{n}\right)^j \left(1 - \frac{1}{n}\right)^{m-j}, \qquad 0 \leqslant j \leqslant m \tag{B.3}$$

The results we derive below are based on an asymptotic argument: what happens when the organization size becomes very big (i.e. $n \to \infty$). We note that the exact calculations in Appendix A yield numbers which very closely fit with the expression we derive below under similar assumptions even when the value of n is as small as 9 or 10. This indicates that what is exactly true at the asymptote is still a pretty good approximation of a large proportion of the overall domain space. To perform an asymptotic proof, we need to distinguish between things that go to infinity at the same rate as n, things that go to infinity at any other rates, and things that stay constant and therefore become negligible compared to n. However, when we use the asymptotic result as an approximation for finite values of n, we need to be careful about what is negligible compared to the values we calculate. Hence, variables

that need to be proportional up to a constant factor in order for a result to hold are treated as if they need to be equal when we interpret the result in the body of the paper.

We distinguish between the following three cases:

- $\frac{m}{n} \to \lambda$ as $n \to \infty$

- $m \ll \sqrt{n}$

- $m \sim n^p$ as $n \to \infty$ for $\frac{1}{2} < p < 1$

B.1.   If $\frac{m}{n} \to \lambda$(constant) as $n \to \infty$

The first result we derive is for when the number of rows grows at the same rate as the number of columns. The argument is identical regardless of whether the row count m is exactly equal to n (in the case where all individuals perform their own independent rating) or whether it is proportional to n.

Theorem 1

$$\text{If} \quad \frac{m}{n} \quad \to \quad \lambda > 0,$$

$$\text{then} \quad \frac{M_n}{a_n} \quad \to \quad 1 \quad \text{as} \quad n \to \infty,$$

where $a_n = \frac{\ln(n)}{\ln(\ln(n))}$.

Proof:

The proof of Theorem 1 consists in showing that:

$$P[M_n \geqslant k_n^+] \to 0 \text{as } n \to \infty \tag{B.4}$$

$$\text{and}$$

$$P[M_n < k_n^-] \to 0 \text{as } n \to \infty \tag{B.5}$$

$$\tag{B.6}$$

where for $\epsilon > 0$, $k_n^+ = \lfloor a_n (1 + \epsilon) \rfloor$ and $k_n^- = \lfloor a_n (1 - \epsilon) \rfloor$.

The first case is straightforward:

$$
\begin{aligned}
P\left[M_n \geqslant k\right] &\leqslant \sum_{i=1}^{n} P\left[C_i \geqslant k\right] \\
&\leqslant n P\left[C_1 \geqslant k\right] \\
&= n \sum_{j=k}^{m} \binom{m}{j} \left(\frac{1}{n}\right)^j \left(1 - \frac{1}{n}\right)^{m-j} \\
&\leqslant n \sum_{j=k}^{m} \frac{1}{j!} \left(\frac{m}{n}\right)^j \\
&= \frac{n}{k!} \left(\frac{m}{n}\right)^k \sum_{j=k}^{m} \frac{k!}{j!} \left(\frac{m}{n}\right)^{j-k} \\
&= \frac{n}{k!} \left(\frac{m}{n}\right)^k \sum_{j=k}^{m} \frac{1}{(k+1)\ldots(j)} \left(\frac{m}{n}\right)^{j-k} \\
&\leqslant \frac{n}{k!} \left(\frac{m}{n}\right)^k \sum_{j=0}^{\infty} \frac{1}{(k+1)^j} \left(\frac{m}{n}\right)^{j} \\
&= \frac{n}{k!} \left(\frac{m}{n}\right)^k \left(1 - \frac{m}{(k+1)n}\right)^{-1}.
\end{aligned}
$$

We know that $\frac{m}{n}$ converges to the constant λ, so if we use a k which continues to grow (such as the k_n^+ we defined above,) we will get an upper bound on $P\left[M_n \geqslant k_n^+\right]$ given by :

$$
P\left[M_n \geqslant k_n^+\right] \quad \leqslant \quad c \, \frac{n}{k_n^+!} \, (2\lambda)^{k_n^+} .
$$

for any constant c we choose. It remains to show that the right hand side of the expression converges to zero as $n \to \infty$, which we do by taking a logarithm and showing that it goes to $-\infty$, as follows:

$$
\ln\left(\frac{n}{k_n^+!} \, (2\lambda)^{k_n^+}\right) \quad = \quad \ln(n) + k_n^+ \ln(2\lambda) - \ln((k_n^+)!) \tag{B.7}
$$

We know that $\ln(\lambda)$ is a constant, and Stirling's formula (Feller, 1968, Vol. 1, Pp. 52-53) tells us that:

$$
n! \approx \sqrt{2\pi} \, n^{n+\frac{1}{2}} \, e^{-n}. \tag{B.8}
$$

Hence, $\ln\left((k_n^+)!\right) \sim k_n^+ \ln\left(k_n^+\right)$ (where $a_n \sim b_n$ means that $\frac{a_n}{b_n} \to 1$ as $n \to \infty$.)

Since $\ln\left(k_n^+\right)$ grows faster than any constant, we can drop the middle term of equation (B.7) before we expand k_n^+ and drop the lower order terms from the expansion:

$$
\begin{aligned}
\ln\left(\frac{n}{(k_n^+)!}\lambda^{k_n^+}\right) &\sim \ln(n) - k_n^+\ln(k_n^+) \\
&\sim \ln(n) - (1+\epsilon)\frac{\ln(n)}{\ln(\ln(n))}\left(\ln(1+\epsilon) + \ln(\ln(n)) - \ln(\ln(\ln(n)))\right) \\
&= \ln(n) - (1+\epsilon)\ln(n) + o(\ln(n)) \\
&\sim -\epsilon\ln(n)
\end{aligned}
$$

which converges to $-\infty$. (We use the notation $a_n = o(b_n)$ if $\frac{a_n}{b_n} \to 0$ as $n \to \infty$.)

This proves that $P\left[M_n \geqslant k_n^+\right] \to 0$ as $n \to \infty$, so it remains to show that $P\left[M_n < k_n^-\right] \to 0$. This part of the proof depends on the observation that for large n, we can treat the number of votes per column C_k as a set of n independent Poisson variables with parameter 1, as long as we stipulate that the sum of all these variables (i.e. the total number of votes given) is equal to n. In other words,

Lemma 1

$$
P\left[C_1 = k_1, \dots, C_n = k_n\right] = P\left[\mathfrak{N}_1 = k_1, \dots, \mathfrak{N}_n = k_n \,\middle|\, S_n = n\right]
$$

where $\forall\, i, \quad \mathfrak{N}_i$ is Poisson(1), $\mathfrak{N}_i$ independent, identically distributed, and $S_n = \sum\limits_{i=1}^{n} \mathfrak{N}_i$.

Lemma 1 follows from the definition of Poisson distributions. The condition on the total number of votes in the sum of all columns makes the approximation exact at large values of n. Using Lemma 1 on the expression in (B.5):

$$
\begin{aligned}
P\left[M_n < k_n^-\right] &= P\left[\mathfrak{M}_n < k_n^- \,\middle|\, S_n = m\right] \\
&= \frac{P\left[\mathfrak{M}_n < k_n^-\right] P\left[S_n = m \,\middle|\, \mathfrak{M}_n < k_n^-\right]}{P\left[S_n = m\right]}
\end{aligned}
\qquad\text{(B.9)}
$$

where $\mathfrak{M}_n = \max\left(\mathfrak{N}_1(\lambda), \mathfrak{N}_2(\lambda), \dots, \mathfrak{N}_n(\lambda)\right)$, the maximum of n Poisson random variables with parameter $\lambda = \frac{m}{n}$, and S_n is as defined in Lemma 1.

We now prove that (B.9) converges to zero in two steps: First we show that $P\left[\mathfrak{M}_n < k_n^-\right]$ converges to zero as $n \to \infty$, and then we show that the remainder of the expression remains bounded as $n \to \infty$.

$$
\begin{aligned}
P\left[\mathfrak{M}_n < k_n^-\right] &= \left(P\left[\mathfrak{N}_1(\lambda) < k_n^-\right]\right)^n \\
&= \left(1 - P\left[\mathfrak{N}_1(\lambda) \geqslant k_n^-\right]\right)^n \\
\ln\left(P\left[\mathfrak{M}_n < k_n^-\right]\right) &= n\ln\left(1 - P\left[\mathfrak{N}_1(\lambda) \geqslant k_n^-\right]\right) \\
&\leqslant -n\,P\left[\mathfrak{N}_1(\lambda) \geqslant k_n^-\right] \\
\\
&\leqslant -n\,P\left[\mathfrak{N}_1(\lambda) = k_n^-\right] \\
&= -n\,e^{-\lambda}\frac{\lambda^{k_n^-}}{k_n^-!}
\end{aligned}
$$

(by the basic definition of a Poisson distribution.)

Since λ is a constant, and $\ln(0) = -\infty$, showing that $P\left[\mathfrak{M}_n < k_n^-\right] \to 0$ is equivalent to showing that $\frac{n\lambda^{k_n^-}}{k_n^-!} \to \infty$. We do this by taking a second logarithm and using Stirling's formula (B.8):

$$
\begin{aligned}
\ln\left(\frac{n\,\lambda^{k_n^-}}{k_n^-!}\right) &\sim \ln(n) + k_n^- \ln(\lambda) - k_n^- \ln(k_n^-) \\
&= \ln(n) - (1-\epsilon)\ln(n) + o\left(\ln(n)\right) \\
&\sim \epsilon\,\ln(n).
\end{aligned}
$$

Thus, $\frac{n\lambda^{k_n^-}}{k_n^-!} \to \infty$ as $n \to \infty$, which implies that $\ln\left(P\left[\mathfrak{M}_n < k_n^-\right]\right) \to -\infty$, and therefore $P\left[\mathfrak{M}_n < k_n^-\right] \to 0$.

The final step is to show that the remainder of the expression (B.9) remains bounded as $n \to \infty$:

$$
P\left[S_n = m \middle| \mathfrak{M}_n < k_n^-\right] = P\left[N_{1n} + N_{2n} + \ldots + N_{nn} = m\right]
$$

where $N_{11n}, \ldots, N_{nn}$ are i.i.d. random variables with a probability mass function defined as

$$
P\left[N_{in} = j\right] = P\left[\mathfrak{N}_i = j \middle| \mathfrak{N}_i < k_n^-\right]
$$

The Dominated Convergence theorem tells us that:

$$
\begin{aligned}
\mu_n = E(N_{1n}) &\sim E(\mathfrak{N}_1) \\
\sigma_n^2 = var(N_{1n}) &\sim var(\mathfrak{N}_1) \\
x_n = E((|N_{1n} - \mu_n|)^3) &\sim E((|\mathfrak{N}_n - E(\mathfrak{N}_n)|)^3) \\
&\quad \text{as } n \to \infty
\end{aligned}
$$

And by the Berry-Esseen theorem (Feller, 1968, Vol. 2, p. 254):

$$P\left[N_{1n} + N_{2n} + \ldots + N_{nn} = m\right]$$
$$\leqslant P\left[\frac{m-1-n\,\mu_n}{\sqrt{n}\,\sigma_n} \leqslant \frac{N_{1n} + \ldots + N_{nn} - n\,mu_n}{\sqrt{n}\,\sigma_n} \leqslant \frac{m+1-n\,\mu_n}{\sqrt{n}\,\sigma_n}\right]$$
$$\leqslant P\left[\frac{m-1-n\,\mu_n}{\sqrt{n}\,\sigma_n} \leqslant \mathbf{N}(0,1) \leqslant \frac{m+1-n\,\mu_n}{\sqrt{n}\,\sigma_n}\right] + \frac{6\,x_n}{\sqrt{n}\,\sigma_n^3}$$

where $\mathbf{N}(0,1)$ denotes a normal random variable with mean 0 and variance 1. Now if we integrate the probability density function of $\mathbf{N}(0,1)$ between the limits $\frac{m-1-n\,\mu_n}{\sqrt{n}\,\sigma_n}$ and $\frac{m+1-n\,\mu_n}{\sqrt{n}\,\sigma_n}$, we find that

$$P\left[\frac{m-1-n\,\mu_n}{\sqrt{n}\,\sigma_n} \leqslant \mathbf{N}(0,1) \leqslant \frac{m+1-n\,\mu_n}{\sqrt{n}\,\sigma_n}\right] \leqslant \frac{2}{\sqrt{n}\,\sigma_n}\sqrt{\frac{1}{2\,\pi}}$$

This means that $P\left[S_n = m \big| \mathfrak{M}_n < k_n^-\right] = O(\frac{1}{\sqrt{n}})$. But

$$P\left[S_n = m\right] = P\left[Poisson(m) = m\right]$$
$$\sim \frac{1}{\sqrt{2\,\pi\,m}} \quad \text{(by the local central limit theorem.)}$$

So the non-zero portion of (B.9) consists of the ratio of two expressions both of order $1/\sqrt{n}$, with the numerator possibly smaller, so the expression does not diverge as $n \to \infty$, and thus $P\left[\mathfrak{M}_n < k_n^-\right] \to 0$, which completes the proof.

∎

B.2.   If $m \ll \sqrt{n}$

Proposition 1 *If the number of criteria m is a constant and the number of people is allowed to grow, then the column sum will be converge to zero:*

Proof: This follows from Equation (B.3) above:

$$P\left[C_k = 0\right] = \left(1 - \frac{1}{n}\right)^m$$
$$\ln\left(P\left[C_k = 0\right]\right) = m\ln\left(1 - \frac{1}{n}\right)$$
$$\approx -\frac{m}{n} \to 0$$
$$\therefore P\left[C_k = 1\right] \to 0 \quad \text{as } n \to \infty$$

∎

The expected value of the maximal column sum will clearly be 1 in that case because the probability of a column containing more than a single "1" entry becomes miniscule. In fact, that value will be 1 even if the number of criteria grows at a rate proportional to the square root of the number of people:

Proposition 2

$$M_n \to 1 \ as \ n \to \infty \qquad \Longleftrightarrow \qquad \frac{m}{\sqrt{n}} \to 0 \ as \ n \to \infty$$

Proof: Let $Y_1, Y_2, \ldots, Y_m$ be i.i.d. variable from a multinomial distribution $(1; \frac{1}{n}, \frac{1}{n}, \ldots \frac{1}{n})$. Then,

$$P\left[M_n = 1\right] = P\left[Y_1, Y_2, \ldots, Y_m \quad \text{are all distinct}\right]$$
$$= \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{m-1}{n}\right)$$

Hence, $$0 \geqslant \ln\left(P\left[M_n = 1\right]\right) = \sum_{k=0}^{m-1} \ln\left(1 - \frac{k}{n}\right)$$

Using the fact $\ln(1-x) = -x + o(x)$,

$$\ln\left(P\left[M_n = 1\right]\right) \cong -\sum_{k=0}^{m-1} \frac{k}{n}$$
$$= \frac{m(m-1)}{2n}$$

It follows that for large n, the right hand side will tend to zero if and only if $m < \sqrt{n}$, and it will tend to some negative value if $m \geqslant \sqrt{n}$. When the right hand side is 0, then $P\left[M_n = 1\right]$ is 1. ∎

B.3. **If $m \sim n^p$ as $n \to \infty$ for $\frac{1}{2} < p < 1$**

Theorem 2 *Suppose that, as $n \to \infty$, the ratio $m/n^p \to a > 0$, a positive constant. Then,*

$$M_n \to \left\lfloor \frac{1}{1-p} \right\rfloor$$

Proof: As in Theorem 1, for any constant $k \geqslant 0$,

$$P[M_n \geqslant k] \leqslant n \sum_{j=k}^{m} \frac{1}{j!} \left(\frac{m}{n}\right)^j$$

$$\leqslant n \left(\frac{m}{n}\right)^k \sum_{l=0}^{\infty} \frac{1}{(k+l)!} \left(\frac{m}{n}\right)^l$$

$$\leqslant n \left(\frac{m}{n}\right)^k \sum_{l=0}^{\infty} \frac{1}{l!} \left(\frac{m}{n}\right)^l$$

$$= n \left(\frac{m}{n}\right)^k \exp\left(\frac{m}{n}\right)$$

$$\leqslant n \left(\frac{m}{n}\right)^k \text{ By the assumed } m/n \text{ ratio, } n \left(\frac{m}{n}\right)^k \quad \sim a^k n^{1-(1-p)k}.$$

So $P[M_n \geqslant k] \to 0$ provided $1 - (1-p)^k < 0$, i.e. $k \geqslant 1 + \left\lfloor \frac{1}{1-p} \right\rfloor$. This proves that the (integer) value of M_n will have a very small probability of being greater than the value that theorem 2 says it will assume. It remains to show that the probability of M_n being *smaller* than that value is also very small and tend to 0 as $n \to \infty$. This can be shown to follow from Lemma 1 above:

$$P[M_n \leqslant k] = P[\mathfrak{M}_n \leqslant k | S_n = m]$$

$$= \frac{P[\mathfrak{M}_n \leqslant k, S_n = m]}{P[S_n = m]}$$

$$\leqslant \frac{P[\mathfrak{M}_n \leqslant k]}{P[S_n = m]} \tag{B.10}$$

where $\mathfrak{M}_n$ is defined as in equation (B.9) and S_n is as defined in Lemma 1. We need to show that the expression (B.10) tends to 0 as $n \to \infty$ when $k = \left\lfloor \frac{1}{1-p} \right\rfloor - 1$. We examine the numerator and the denominator separately:

$$P[\mathfrak{M}_n \leqslant k] = \left(P\left[\mathfrak{N}_1(\tfrac{m}{n}) \leqslant k\right]\right)^n$$

$$= \left(1 - P\left[\mathfrak{N}_1(\tfrac{m}{n}) \geqslant k+1\right]\right)^n$$

so

$$\ln\left(P[\mathfrak{M}_n \leqslant k]\right) \leqslant -nP\left[\mathfrak{N}_1(\tfrac{m}{n}) \geqslant k+1\right]$$

$$\leqslant -nP\left[\mathfrak{N}_1(\tfrac{m}{n}) = k+1\right]$$

$$= -n\, e^{-\frac{m}{n}} \frac{\left(\frac{m}{n}\right)^{k+1}}{(k+1)!}$$

$$= e^{-\frac{m}{n}} n \left(2a\, n^{p-1}\right)^{\left\lfloor \frac{1}{1-p} \right\rfloor} \frac{1}{\left\lfloor \frac{1}{1-p} \right\rfloor!}.$$

As $n \to \infty$, we see that $P[\mathfrak{M}_n \leqslant k] \leqslant e^{-cn}$ for some constant c.

The proof follows from noting that e^{-cn} converges to zero much faster than the expression for the denominator of expression (B.10), which we show below:

$$P\left[S_n = m\right] = P\left[Poisson(m) = m\right]$$

$$\sim \frac{1}{\sqrt{2\pi m}} \quad \text{(by the local central limit theorem)}$$

Recalling that $m = an^p$,

$$P\left[S_n = m\right] = O(n^{-p/2})$$

as $n \to \infty$, which, as we wish to prove, is smaller than e^{-cn}.

■

Notes

1. The authors anticipate engaging in several such studies in the near future. On such study will be reported in a forthcoming paper in this Journal.
2. Using any linear combination of several list rankings will also yield the same results as long as *either* n is very large, *or* rankings may take on any value on the real number line.
3. The results of this theorem apply under certain conditions on the distribution of d_n, which are described in the reference. For modeling convenience, we assume that these conditions are satisfied.
4. This form of the solution was contributed by Mark Scandera, PhD candidate at the Department of Mathematics, Massachusetts Institute of Technology *skan@math.mit.edu*

References

Banks, D. L., and K. M. Carley. 1996. Models for network evolution. *Journal of Mathematical Sociology* 21(1–2):173–196.

Brass, D. J., and M. E. Burkhardt. 1992. Centrality and power in organizations. In N. Nohria and R. G. Eccles (Eds.), *Networks and Organizations : Structure, Form and Action*. Boston, MA: Harvard Business School Press.

Burt, R. S. 1992. *Structural Holes : The Social Structure of Competition*. Cambridge, MA: HarvardUniversity Press.

Carley, K. M. 1990. Group stability: A socio-cognitive approach. *Advances in Group Processes* 7:1–44.

Carley, K. M. 1991. A theory of group stability. *American Sociological Review* 56:331–354.

DiMaggio, P. 1992. Nadel's paradox revisited : Relational and cultural aspects of organizational structure. In N. Nohria and R. G. Eccles (Eds.), *Networks and Organizations : Structure, Form and Action*. Boston, MA: Harvard Business School Press.

Feller, W. 1968. *An Introduction to Probability Theory and its Applications*. Chichester, New York: John Wiley and sons.

Freeman, L. C. 1997. Uncovering organizational hierarchies. *Computational and Mathematical Organization Theory* 3(1).

Huberman, B. A., and T. Hogg. 1995. Communities of practice. *Computational and Mathematical Organization Theory* 1(1).

Kaufer, D. S., and K. M. Carley. 1993. *Communication at a Distance : the Influence of Print on Sociocutural Organization and Change*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Scott, J. 1991. *Social Network Analysis : A Handbook*. London: Sage Pubications.

Walid Nasrallah is a Ph.D. candidate at the Civil Engineering department at Stanford University. He earned his Masters degree at MIT in 1989, worked in the construction and software industries, then obtained his Engineer's Degree at Stanford in 1996. His research interests include the evolution of organizations in response to new

technologies, and optimization of communication networks in existing organizations.

Dr. Raymond Levitt is Professor of Civil Engineering and Associate Director of the Center for Integrated Facility Engineering, (CIFE) at Stanford University since 1980. He earned a BSCE degree format the University of Witwatersrand, and then worked in construction in South Africa and Canada. He earned MS and Ph.D. degrees in Civil Engineering from Stanford. From 1975-1980, he served on the faculty of MIT's Civil Engineering Department, heading up MIT's Construction Management Program.

Dr. Levitt's research and teaching have focused on decision-making and communication in project teams and companies. Since 1987, his Virtual Design Team (VDT) research group has focused on developing new organization theory and computational models to design optimal work process and organization configurations for concurrent product development teams.

Dr. Levitt has been a consultant on IT and enterprise engineering for numerous projects and companies. In July 1996, he took leave from Stanford University to launch Vité Corporation to commercialize the results of VDT. He currently serves as a Director of Vité and Design Power, Inc.

Peter Glynn received his Ph.D. from Stanford University, after which he joined the faculty of the Department of Industrial Engineering at the University of Wisconsin-Madison. In 1987, he returned to Stanford, where he is the Thomas Ford Professor of Engineering in the Department of Engineering-Economics Systems and Operations Research. He is a co-winner of the 1993 Outstanding Simulation Publications Award sponsored by the INFORMS College on Simulation and is a Fellow of the Institute of Mathematical Statistics. His research interests include discrete-event simulation, computational probability, queuing, and general theory for stochastic systems.