IBM Research Report

Towards simultaneously exploiting structure and outcomes in interaction networks for node ranking

Kashyap Dixit
IBM Research Division
IBM India Research Lab
New Delhi, India

S Kameshwaran
Indian School of Business
Hyderabad, India

Sameep Mehta
IBM Research Division
IBM India Research Lab
New Delhi, India

Vinayaka Pandit
IBM Research Division
IBM India Research Lab
New Delhi, India

N Viswanadham
Indian School of Business
Hyderabad, India

IBM Research Division
Almaden - Austin - Beijing - Delhi - Haifa - T.J. Watson - Tokyo - Zurich

Abstract

In this paper, we present algorithms for ranking nodes in interaction networks. Informally, they capture the patterns of historical interaction among the nodes and the associated outcomes. There exists a cardinal ranking over the set of outcomes, characterizing the order of preference. We argue that ranking of nodes should be influenced by both structural properties of the networks and the outcome/value created by the interactions. The former aspect is well studied in social network analysis and is accounted for, in various measures like centrality, reputation, influence etc. However, the latter aspect is largely unexplored. Our proposed algorithms simultaneously take into account both structural properties as well as the outcomes to assign ranks for the nodes. We develop a novel eigenvector-like computation that exploits the structural influences, importance of value creation, and any exogenous information available to the ranking system. We report experimental results on the IMDB dataset.

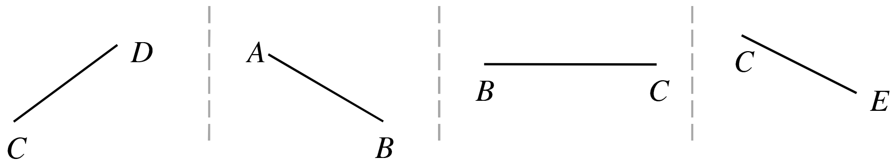
1 Introduction

In most real world scenarios, interactions involves a set of actors[1] working towards a common goal. Examples of such interactions involve software developers working on a project delivery, actors working in same movies, researchers collaborating for papers, service providers executing supply orders etc. When several of these individual interactions are aggregated, a static network view of the past interactions is obtained. We refer to such networks as *interaction networks*. The structural and behavioral properties of the network and its constituent actors can be studied using various (social) network analysis techniques. Social network literature is abundant with networks related to interactions such as telephone calls, e-mails, friendship, club membership, co-authorship, business partnership, etc. Ranking of the actors in the resulting network has been of research interest since the inception of social network analysis. The algorithms for ranking vary depending on the objective, like measuring *authority*, *influence*, *status*, *prestige*, and on the underlying community represented by the network. All these properties exploit the structural (connectivity) properties of the network. In this paper, we introduce the notion of *value creation* for the interaction networks. The readers would agree that in the professional communities individual interactions result in value creation for the system. Ranking the entities solely on structural properties misses this dimension and may result in a ranking which is inconsistent with the overall goal of the organization. Similarly, ranking based only on value will overlook the interaction part of the network and will generate a ranking which will be oblivious to the actual interaction dynamics. *We contend that the ranking algorithms for such systems should simultaneously take into account both structural properties as well as value creation aspects.* In this paper, we present such a ranking technique. The proposed interaction network not only captures the interactions among the actors but also encodes the value created (outcome) by the interactions. This is done by adding *special* nodes for each possible outcome and by adding connection between actors and outcomes *appropriately*. For example, interaction network for IMDB data will have nodes for movie actors as well as nodes corresponding to different possible (quantized) user ratings. The edges of the network will depend on the actor lists for different movies. Special edges have to created between nodes corresponding to outcomes and nodes corresponding to the actors.

In this work we consider three aspects which influence the rank of an agent i) number of past interactions, ii) rank of the other agents with whom he interacted, and iii) on the outcomes of the interactions. The above three criteria are not mutually independent and in most communities, they reinforce each other. For example, in the academic and research community of a certain discipline, a researcher's rank is implicitly governed by his past and present collaborators, and also by the outcomes of the collaborations like publications and patents. In the following, we briefly present few communities that are in the interest domain of this paper. As an illustrative example, consider five agents A, B, C, D, E, with four different interactions, as shown in Figure 1. Aggregation of these interactions as an interaction network (assuming for now such an aggregation is meaningful) is shown in Figure 2. With this example, we present possible outcomes in three different professional communities. Typically, the aggregate graph shown in Figure 2 is used in the traditional social network analysis. The main contribution of this paper is to show that an augmented graph that reflects both structure and outcomes offers a very promising approach to rank the nodes.

[1]In this paper, we use actors, entities, nodes and agents interchangeably.

Individual Interactions



Illustrative Outcomes for Different Communities

	Service Interactions				
I.	Deal acquisition *Success*(S) / *Failure*(F)	*S*	*S*	*F*	*S*
II.	Revenue generated (in millions of USD)	28	32	15	17
	Co-authorship				
III.	Journal impact factor	1.409	2.143	1.102	0.326
IV.	Citation count	32	12	15	47
	Movie co-stars				
V.	IMDB ratings (Out of 10)	6.7	7.4	8.3	9.4
VI.	Box-office revenue (In USD)	55,808,744	70,098,308	100,003,359	107,928,162

Figure 1: Individual interactions with illustrative outcomes from different communities

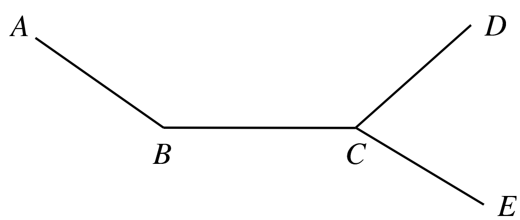


Figure 2: Interaction network of interactions from figure 1

1.1 Service Interaction Networks

One of the distinguishing features of service industry is the high emphasis on people interacting with other people in creating value and serving customers, rather than transforming physical goods [6]. The work presented in this paper was primarily motivated while studying service industry from the perspective of interactions [9]. Let the interactions in figure 1 refer to different deal acquisitions of a software firm. Agents C and D had interacted for the first deal, agents A and B for the second, and so on. The agents can refer to individual employees or teams. The outcome of the interactions is binary: *Success* (S), if the deal was acquired, and *failure* (F), otherwise. Consider another example in service interaction from the construction industry. One of the measures of the value created is the revenue obtained from the execution of a project. Clearly, ranking of the agents in the above two cases should take into account the interactions and the outcomes. Such ranking and associated analysis can be used by the organization in forming effective and efficient teams.

1.2 Co-authorship Networks

In the academic and research community, people collaborate with others in publishing articles, which is a predominantly used measure in assessing the status of a researcher. Let the agents in the above example be researchers, with each interaction resulting in a journal publication. The interaction network of figure 2 is a *co-authorship* network, which is well studied in the literature. Ranking in such a network can only take into account the number of publications and the *popularity* of the co-authors. However, in reality, the value created out of the publications play an important role in a researcher's status. In figure 1, two different measures are given to assess the value created out of individual interactions: *journal impact factor* and *citation count.* For an individual researcher, a measure such as Erdös number captures the structural properties of network whereas h index [8] captures the citation impact of the publications. One can be easily convinced that ranking in co-authorship networks should take into account both measures to generate a realistic and acceptable ranking.

1.3 Movie co-stars

In the movie industry, popularity of the stars to a large extent depend on their co-stars' popularity (at least initially). This, in turn, is based on the popularity of movie and vice-versa. An unknown star can become popular by acting in a hit movie with a popular star. As mentioned earlier, the criteria number of interactions, status of interacting agents, and outcome of interactions reinforce each other. In the example, each interaction refers to a movie and the possible measures of value creations are IMDB ratings and box office revenues. We consider this example in more detail later, where we apply our proposed ranking methodology in ranking movie stars based on the ratings received at IMDB.

1.4 Our work

We develop a formal model for interaction networks with outcomes and define the problem of ranking nodes in this context. We then develop an algorithm based on the *alpha centrality* measure introduced by Bonacich [2]. The measure is based on a eigenvector-like computation. Our main insight is on constructing an augmented directed graph with special nodes corresponding to different outcomes. We also show that popularly used Singular Value Decomposition (SVD) doesn't capture all intricacies and therefore, can not be directly applied to problem at hand.

To reiterate, the key contributions of the paper are:

- We introduce the concept of value creation in professional networks and argue why this aspect should play an important role ranking nodes in an interaction network.
- We present a ranking algorithm which takes into account the structural as well as value creation aspect of the interaction networks.
- We present result on real dataset from IMDB.

The rest of the paper is organized as follows. In Section 2.1, we discuss related literature on node ranking in graphs. In Section 2.2, we show how some of the commonly used approaches give unsatisfactory results, even on the simple example considered in Figures 1 and 2. In Section 3.1, we formally define the problem and in Section 3.2, we present our approach to ranking nodes in interaction networks taking both structure and outcome into account. Section 4 presents our experimental setup and results.

2 Node Ranking in Networks

Ranking of nodes in a network with respect to some *importance* measure is an active research topic in the fields of social network analysis, network data mining, and in general complex networks. A node refers to an *entity* that could be an individual, role, group, or an organization. Edges encode relationships between nodes, which are of two kinds: *Persistent* (weblinks, friendship, membership, affiliations) and *discrete interactions* (e-mails, collaboration, authorship, team work). In the following we briefly review some of the works in node ranking relevant to interaction networks.

2.1 Related Literature

Various methods exist for ranking nodes in a network. Degree Centrality is one such measure. The nodes in the network are ranked based on the number of other nodes to which they are connected. In interaction networks, this translates into *experience* or *number of interactions*: Rank the agents based on the total number of interactions. Eigenvector centrality [2] is a popular measure for ranking in social network analysis. Eigenvector centrality measures the centrality of a node as a linear combination of centralities of nodes to which it is connected. Unlike degree, which weighs every adjacent node equally, the eigenvector weights adjacent nodes according to their centralities. Let $A = [a_{ij}]_{N \times N}$ be the (weighted) adjacency matrix of a network with N nodes. The measure x_i for node i depends on the status measures of the interacting nodes $x_i \propto \sum_j a_{ij} x_j$. This can be expressed in matrix notation as $\lambda x = Ax$. Here, λ is the largest eigenvalue and x is the corresponding eigenvector. The agents can then be ranked based on the component wise value of the eigenvector x. The idea of using the eigenvector to do ranking dates back to the 1950's [16, 10], which was also later used by HITS and PageRank to rank pages in the web [15, 11]. While there are various extensions to HITS and PageRank, we will not be surveying those. The eigenvector measure is very relevant to the interaction networks: Status of an agent depends on the status of the agents with which it interacts. An author's popularity is affected by the popularity of the co-author. Also, it takes into account the degree (the weights a_{ij}) of the nodes implicitly. Co-authorship networks are an important class of networks and have been used extensively to determine the status of individual researchers and the structure of scientific collaboration [13, 14]. To rank the authors in a co-authorship network, *AuthorRank* algorithm was proposed in [12]. It is basically a modification of *PageRank* [15] algorithm for a weighted, directional network. PageRank is originally designed to rank retrieval results based on the hyperlink structure (persistent relationship) of the web, which is directed but binary graph. The ranking is similar to eigenvector ranking where a page has high rank if the sum of the ranks of its *backlinks* is high. In AuthorRank, the construction of the weights on the links between the authors is different from that of the PageRank. The co-authorship frequency and the number of authors per publication were taken into account in the network construction. Some context aware measure have also been proposed. Haveliwala [7] developed topic-sensitive PageRanks by computing several eigenvectors biased by specific topics. The rankings for a query was obtained by appropriate combination of precomputed eigenvectors. White and Smyth [17] developed an algorithm to rank the relative importance of nodes in a network with respect to a set of "root nodes". Their technique is based on the notion of weighted paths. A node is considered to be important to a root node if they share many short paths. Chitrapura and Kashyap [4] presented a flow based algorithm to rank webpages based on their relevance to the user query. Delong et al. [5] proposed design the webgraph such that it captures both the link information as well the underlying concepts of the webpages. Asur et al [1] presented algorithms to find important nodes in dynamic networks by studying the node evolution and interactions.

2.2 Ranking based on Structure and Outcomes of Interactions

The above ranking schemes when applied to interaction network will only take into account the structural properties. But, as stated, our goal is to develop a ranking scheme that takes structural properties and outcomes of interactions. Consider the interaction network of Figure 2 arising out of interaction from figure 1 with outcomes *II* (revenue generated) from the domain of service interactions. Let the weight on each edge be unity. We will use this example to highlight some of the deficiencies of past approaches. We will begin with an ideal ranking that takes both structure and outcomes into account. The results of the previous ranking methods is tabulated in 1.

2.2.1 Expected Ranking based on Structure and Outcomes of Interactions

Consider the network in Figure 2 with outcomes *II* (refer Figure 1). It is obvious that C should be ranked first, followed by B as both have high number interactions and thus their share of outcomes. D and E are same with respect to their interactions, but D has generated more revenue than E, hence should be ranked ahead of E. A has slightly more value created than D, but their interaction strengths are opposite: A's partner B is ranked less than D's partner C. As there is no large difference in the revenue generated by D and A, we can expect to rank D ahead of A. On the contrary, the difference in revenue generation between A and E is quite large and it will counteract the E's reputation of having worked with C. Putting all together, we can expect the ideal ranking to be C, B, D, A, E.

Let us now consider the rankings obtained by the past approaches.

2.2.2 Degree Ranking

Degree ranking merely counts the number of interactions and hence quantifies only the experience. The ranking due to degree is: $C, B, \{A, D, E\}$.

2.2.3 Eigenvector Ranking

Eigenvector is the most appropriate ranking technique for analyzing the structure of interactions. It models the *inherited* or *transfered* status and implicitly takes into account the degree ranking. Let A be the adjacency matrix of the network in figure 2. The largest eigenvalue is 1.8477 and the corresponding eigenvector is 0.271 (A), 0.5 (B), 0.653 (C), 0.354 (D), and 0.354 (E). The ranking is:

Methodology	Ranking
Degree	$C, B, \{A, D, E\}$
Eigenvector	$C, B, \{D, E\}, A$
Outcome	C, B, A, D, E
Expected (Structure + Outcomes)	C, B, D, A, E

Table 1: Rankings based on different methodologies

$C, B, \{D, E\}, A$. It is obvious that C is ranked first, followed by B, and A is ranked behind $\{D, E\}$ (as $\{D, E\}$ is connected C, which has higher ranking than B to which A is connected). Although this ranking is quite satisfactory for this example, its main deficiency is the failure take outcomes into account.

2.2.4 Outcome based Ranking

Each of the individual interactions has resulted in different revenues. Let each interacting agent has equal contribution to the outcome for an interaction. Thus, for the first interaction, C and D has contributed USD 10 million to a total revenue USD 20 million. Considering only the value created from the interactions, one can obtain ranking of agents. For this example, the total value created by the agents are: 16 (A), 23.5 (B), 30 (C), 14 (D), and 8.5 (E). The resulting ranking is: C, B, A, D, E. This method fails to take the structural aspect into account.

From the above discussion, we observe that the intended ranking mechanism should be similar in spirit to eigenvector ranking and in addition should consider the outcome values. In the next section, we describe the problem in detail followed by the proposed methodology.

3 Model

3.1 Problem Definition

Let $\{1, 2, \dots, N\}$ be the set of agents of interest and $\{1, 2, \dots, T\}$ interactions among these agents have been observed. An interaction results in one of the $\{1, 2, \dots, M\}$ outcomes. Following indices are used: i and j for agents, t for interactions, and m for the outcomes. For sake of brevity, an index when used with the $\forall$ notation denote the corresponding universal, unless specified otherwise. For example, $\forall m$ is used instead of $\forall m \in \{1, 2, \dots, M\}$.

Let $\rho_m \in \mathcal{R}$ denote the utility of outcome $m \in \{1, 2, \dots, M\}$. The $\{\rho_m\}$ can be cardinally ordered and if $\rho_{m'} > \rho_{m''}$, then the outcome m' is preferable to m''. The outcome I of figure 1 is of categorical type S and F. The utility measures have to given for such outcomes. For outcome II, the value created can be directly used as utilities or an utility function has to be defined. As the outcomes are finite, the possible utility values should also be finite. For example, the IMDB viewers' ratings (V) can be appropriately quantized to have 10 or 20 discrete outcomes.

An interaction $t \in \{1, 2, \dots, T\}$ involves a subset of agents $V_t \subseteq \{1, 2, \dots, N\}$. The pattern of the interaction is given by the edge set E_t with a non-negative δ_{ij}^t denoting the weight on the edge $(i, j) \in E_t$. The weight for non-existent edges is zero: $\delta_{ij}^t = 0$, $(i, j) \ni E_t$. As t is an interaction among the V_t agents, the graph (V_t, E_t) is connected. The specific structure of the graph is given by the nature of the interaction. If the interaction is that of a *group* work with one interaction involving all, then the graph is complete. For an hierarchical interaction, the graph will have a tree structure. Let $R_t \in \{1, 2, \dots, M\}$ be the observed outcome of the interaction. The interaction t can thus be completely characterized by the graph $(V_t, E_t, \{\delta_{ij}^t\})$ and the outcome R_t.

Given the past t interactions $\{(V_t, E_t, \{\delta_{ij}^t\}, R_t)\}$ and the outcome utilities $\{\rho_m\}$, the problem is to rank the agents $\{1, 2, \dots, N\}$ according to the measure that takes into account the inter-agent interactions and the outcomes of the interactions. As mentioned in previous section, ranking based on inter-agent interactions can be done using eigenvector centrality. In the following, we propose the methodology for including the outcomes in the ranking.

3.2 Methodology

Firstly, the individual interactions are aggregated to obtain the *agent interaction network*, which, with a slight abuse of notation, is given by the graph $(V, E, \{\delta_{ij}\})$:

$$V = \cup_t V_t \quad (1)$$
$$E = \cup_t E_t \quad (2)$$
$$\delta_{ij} = \sum_t \mu^t \delta_{ij}^t \ \forall i, j \quad (3)$$

The overall strength of interaction between any two agents is the linear combination of the individual strength of interactions. If the interactions are chronologically ordered, then μ_t can be used as the *past influence factor* to model the relative importance of the interactions with respect to time. By judiciously choosing $\{\mu^t\}$, one can model various kinds of past influences: *Uniform*, *sliding window*, etc. We call the edge weight matrix $A = [\delta_{ij}]$ of order N as the *agent interaction matrix.* The matrix A captures the past interactions between the agents and outcomes of the interactions are not considered. In the simplest construction, it will be a weighted adjacency matrix. In order to take into account the outcome of interactions, we augment the outcomes as nodes to the interaction network $(V, E, \{\delta_{ij}\})$ and the resulting graph is directed.

3.2.1 Augmenting the Outcomes as Nodes

Let the outcome of interaction t be m. The graph $(V_t, E_t, \{\delta_{ij}^t\})$ is updated as follows:

$$V_t \leftarrow V_t \cup \{m\} \quad (4)$$
$$E_t \leftarrow E_t \cup \{(m, i)\}, \ \forall i \in V_t \quad (5)$$
$$\delta_{mi}^t \geq 0, \ \forall i \in V_t \quad (6)$$
$$\sum_{i \in V_t} \delta_{mi}^t = 1 \quad (7)$$

The outcome m is added as a node and a directed edge is added from m to each of the other agents that participated in the interaction. The weights on the newly added edges that are given by (6) and (7), captures the relative contribution of agents in realizing the outcome m. The weights on the non-existent edges are zero: $\delta_{im}^t = 0, \forall i$ and $\delta_{mi}^t = 0, \forall i \notin V_t$.

Without loss of generality, we can assume that each of the outcomes is realized in at least one of the interactions. The aggregation of the graphs augmented with outcomes is given by:

$$V \leftarrow V \cup \{1, 2, \ldots, M\} \quad (8)$$
$$E \leftarrow E \cup \{(m, i) : \exists t, \ R_t = m \wedge i \in V_t\} \quad (9)$$
$$\delta_{mi} = \sum_t \omega^t \delta_{mi}^t, \ \forall i, m \quad (10)$$
$$\delta_{im} = 0 \quad (11)$$

The outcomes $\{1, 2, \ldots, M\}$ are added as nodes to the interaction network and an edge from outcome m to a node i exists if i had been a part of at least one interaction with outcome m. The weights on the outcome-agent edges are are taken as linear combination of the corresponding weights in the individual interactions. We call the above network as the *agent-outcome interaction network* and the corresponding edge weight matrix as the *agent-outcome interaction matrix* Δ of order $(N + M)$. For the network in figure 2 with outcome *II* of figure 1, the agent-outcome interaction network is shown in figure 3. The four outcomes $O1, O2, O3, O4$ correspond to the outcomes of the four interactions. Assuming equal contribution from each agent, weight on the directed edge from an outcome to an agent is $1/2$.

The matrix Δ captures both the inter-agent interactions and agent-outcome interactions. In particular, notice the directed edges from the outcomes to the agents and that there is no edge from the agents to the outcomes. It is to capture the fact that ranking of agents is a function of outcomes, and not vice versa. The outcomes are the result of agent interactions. The overall intended effect of the directed construction is to let the outcome-nodes transfer their weights to the agents and not allow any transfer of weights from agents to outcomes. The matrix Δ, however does not take into account the utilities of the outcomes $\{\rho_m\}$. We treat the utilities as exogenous status of the outcomes and combine with Δ to obtain the ranking. The above methodology is motivated by a ranking technique called as *alpha-centrality* [3].

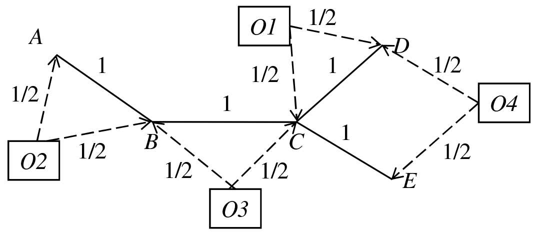


Figure 3: Interaction network with outcomes augmented as nodes

3.2.2 Alpha-Centrality Measure

The eigenvector centrality [2] has a limitation that it can only be applied to non-negative, symmetric matrices. Thus, if the underlying graph is asymmetric, especially with *unchosen* nodes (zero in-degree nodes), the ranking provided by eigenvector centrality is inconsistent. The unchosen nodes receive no status from the other nodes and hence contribute nothing to the nodes to which they are connected. Alpha-centrality measure, circumvents this problem by allowing each node to have status that is independent of its connections to other nodes. This exogenous status vector e provides status for unchosen nodes. Let us consider the agent-outcome interaction network. Clearly, the outcomes are unchosen nodes. The alpha-centrality measure x for matrix Δ is:

$$x = \alpha\Delta^T x + e \tag{12}$$

The first term in the above equation is similar to the eigenvector relation. It was shown in [3] that alpha-centrality measure is similar to that of the eigenvector, when applied to problems where they can be compared. For example, when applied to symmetric relations with the same status for all nodes, both measures are same for $\alpha = 1/\lambda$, where λ is the largest eigenvalue of the connection matrix. As evident from the relation (12), alpha-centrality considers both inter-node connections and the exogenous status of the nodes. The measure x is given by:

$$x = (I - \alpha\Delta^T)^{-1} e \tag{13}$$

The vectors e and x, and identity matrix I are of order $(N + M)$. The suggested value for α is in the range $(0, 1/\lambda)$ [3]. It is worth noting that in eigenvector centrality, the vector corresponding to the largest eigenvalue is chosen as the measure for ranking. In the following, we describe how to choose e and α for our agent-outcome interaction network.

3.2.3 The Exogenous Status Vector e

Let us assume for now that we have chosen $\alpha \in (0, 1/\lambda)$. The vector e for our problem is $(e_1, e_2, \ldots, e_N, e_{N+1}, \ldots, e_{N+M})$, where the first N components are for external statuses of the agents and the last M are for that of the outcomes. The component e_i quantifies the external status of agent i that is independent of Δ and similarly e_m quantifies the external status of outcome m. We define e as follows:

$$e_i = 1, \ \forall i \tag{14}$$

$$e_{N+m} = \theta\rho_m, \ \forall m \tag{15}$$

There is no external status for the agents (as per the problem definition) that can be a criterion for ranking. Hence, all agents are given equal status of unity. The outcomes on the other hand have the utilities as the external status. The utilities are unique up to a linear transformation and hence the external status of an outcome is defined as proportional to its utility by a non-zero scalar θ. In other words, $\{\theta\rho_m\}$ still preserves the cardinal structure and the relative magnitude of the utilities are unchanged.

As mentioned earlier, the external status is supposed to be independent of the interactions. By the above usage, it would appear that the outcome utilities $\{\rho_m\}$ are independent of Δ and hence the interactions. The utilities are result of the interactions and hence the above usage would appear wronged. Our intention is to use the utilities as one of the means to rank the agents in addition to the interactions. We are not interested in explaining the role of an agent or its contribution in obtaining an outcome, but rather simply use the outcomes as a part of a measure of the agents'

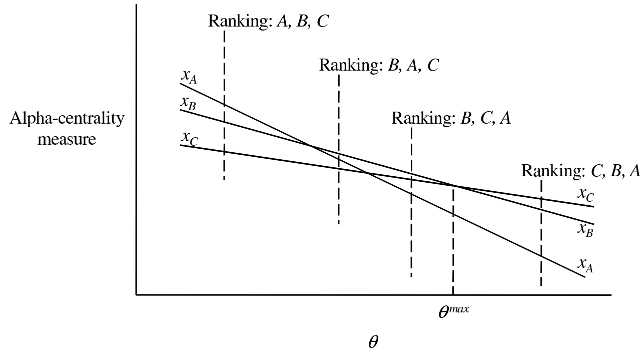


Figure 4: Alpha-centrality measure as a function of θ

α	**Ranking**
0.054, 0.102, 0.15, 0.198	C, B, A, D, E
0.247, 0.295, 0.343	C, B, D, A, E
0.391, 0.439, 0.487	C, B, D, E, A

Table 2: Ranking for different α values

status. Note that outcomes are also nodes in our model and hence the alpha-centrality will provide some measure to all the outcomes. However, as the outcomes are unchosen nodes, their only source of status is their external status and hence *should* be proportional to it. Define matrix Y of order $(N+M)$:

$$Y = (I - \alpha\Delta^T)^{-1} \tag{16}$$

By definition of Δ,

$$y_{N+m,N+m} = 1, \; \forall m \tag{17}$$

$$y_{N+m,N+m'} = 0, \; \forall m, m \neq m' \tag{18}$$

Rephrasing x vector in terms of Y,

$$x_i = \sum_j y_{ij} e_j + \sum_m y_{i,N+m} e_{N+m}, \; \forall i \tag{19}$$

$$x_{N+m} = e_{N+m}, \; \forall m \tag{20}$$

Substituting for the e, as given defined in (14) and (15),

$$x_i = \sum_j y_{ij} + \theta \left(\sum_m y_{i,N+m} \rho_m \right), \; \forall i \tag{21}$$

$$x_{N+m} = \theta \rho_m, \; \forall m \tag{22}$$

Thus the outcome components of x have the same value as their corresponding external status. On the other hand, the $\{x_i\}$ are dependent on θ and rankings could possibly change with respect to the choice of θ. As the other terms in (21) are constants, we have N lines given by equations (21) as a function of θ. If all these lines are parallel, then then ranking is independent of θ. Otherwise, the rankings will depend on θ as shown in figure 4. The ambiguity in ranking is due to the magnitudes of Δ, θ, and $\{\rho_m\}$. The Δ and $\{\rho_m\}$ are instance specific and hence θ should be chosen such that ranking is not affected. One possibility is to choose a positive $\theta > \theta^{max}$ (greater than all the intersecting points, as shown in figure 4), so that the ranking will remain unchanged with further increase in θ value. As each line will possibly intersect at $N-1$ points with $N-1$ lines, the total number of possible intersecting points are $(N-1)(N-2)/2$. The θ^{max} is the maximum of the above intersecting points, which can be estimated in $O(N^2)$ time.

3.2.4 The α Value

Another free parameter that affects the ranking is the $\alpha \in (0, 1/\lambda)$. The α reflects the relative importance of endogenous versus exogenous factors in the determination of centrality [3]. For the network in figure 3, the largest eigenvalue of the Δ matrix is 1.8477 and hence the α has to be chosen in range $(0, 1/1.8477)$. The table 2 shows the rankings obtained for ten different equally spaced α values in the above range. The ranking given by mid-range α values is that of the expected ranking, whereas the low range is same as that of outcome rank and the higher range is more similar to eigenvalue and degree rank (also see table 1). However, only with experimentation on larger networks, we can generalize the above parametric characterization.

3.2.5 Discussion

Consider the matrix Δ obtained from the special construction described in Section 3.2.1. This matrix is an asymmetric matrix. Given the success of eigenvector based ranking for other graphs, one would be tempted to consider the ranking according the left singular vector corresponding to the largest singular value of Δ. However, this naive approach does not seem very productive. The intuition being that the augmented directed network has a large symmetric component corresponding to actor nodes and a small set of directed edges from the outcome nodes to the actor nodes. Let us consider the following thought experiment: Suppose, we change the outcomes a subset of interactions. Ideally, it should affect the ranking of the actor nodes which participated in the subset of interactions whose outcome was changed. Let Δ' be the matrix obtained after the modification of outcomes. The difference between Δ and Δ' is only in the edges from outcomes nodes to actor nodes. Therefore, structurally, they are quite similar. Therefore, SVD based rankings for the two matrices will not be significantly different. This is also shown in our experiments.

4 Experiments

In this section, we present experimental results that show the efficacy of our approach. The experiments were designed in such a way that the results are easy to interpret and verify. The experiments also are geared to bring out the conceptual strength of our approach. Our data source and experiments also make it easy for us to build anecdotal evidences. The experiments are designed from the data extracted from the Internet movie database (IMDB)[2]. IMDB allows us to programmatically extract the following movie lists: list of all movies, list of movies according to genre and time period, within a genre top user rating movies, within a genre a random mix of movies and so on. We also get the user ratings corresponding to the movies in the lists that we work with. Additionally, it is also easy to extract the list of actors appearing in a movie. So, for a given set of movies, we construct the interaction network as follows: each actor who appears in any of the movies is an agent, each movie represents an interaction that is incident on all the actors in its list, and the outcome associated with the interaction is the average user ratings for the movie. Given this interaction data, we construct the Δ matrix as described in Section 3.2.1. Our experimentation explores the effects of the different constructions for the exogenous vector e. Some of the important conceptual aspects that we exhibit from our experiments are:

- The rankings obtained from *alpha centrality* based approach matches with our intuitive knowledge of the domain.
- A purely structural approach like SVD, even on the asymmetric matrix Δ having the outcome nodes, fails to respond to changes in the outcomes.
- When the outcomes associated with a certain subset of movies is changed, the rankings of the actors also undergo corresponding change. This will indicate that our approach takes the outcomes into account when ranking.
- When we increase/decrease the importance of a subset of actors, then, the actors who are structurally better connected to them should also see corresponding change in their importance. This will indicate that our approach takes the structural aspects into consideration as well.
- Relative importance of actors over different periods of times can also be demonstrated.
- The choice of e is important to obtain good results. We demonstrate some effective and some not so effective constructions for e.

[2] `http://www.imdb.com/interfaces`

Actor Index	List 1	List 2
1	Marlon Brando	Marlon Brando
2	Al Pacino	James Mason (I)
3	Robert De Nero	Louis Calhern
4	Sean Bean	Glenn Ford (I)
5	Jean Reno	Karl Malden
6	Don Cheadle	Ben Johnson (I)
7	John Travolta	Timothy Carey
8	Hugh Jackman	Richard Harris (I)
9	George Clooney	Montgomery Clift
10	Casey Affleck	Dean Martin (I)
11	Brad Pitt	Frank Overton
12	Matt Damon	Malcolm Atterbury
13	Dan Fredenburgh	Robert Ryan (I)
14	Bill Nighy	Burt Lancaster
15	Johnny Depp	Frank Sinatra
16	Orlando Bloom	Ernest Borgnine
17	Jack Davenport	Lee Marvin
18	Lee Arenberg	Rhys Williams (I)
19	Tom Hollander	DeForest Kelley
20	Jude Law	John Wayne (I)
21	Anthony Hopkins	Walter Brennan
22	Sean Penn	Ed Wynn
23	Samuel L. Jackson	Stephen Boyd (I)
24	Kevin Bacon	Milton Berle
25	Tom Hanks	Tony Bennett (I)
26	Steve Buscemi	Al Pacino
27	Clive Owen	Robert De Niro
28	Nicolas Cage	Broderick Crawford
29		Ricky Nelson (I)
30		Buddy Ebsen

Table 3: List of actors used in experiments

The IMDB dataset comes in very handy in achieving our objective. One of the advantages is that, relative rankings of actors is somewhat easy to appreciate (considering that movies seem to be an indispensable aspect of contemporary experience!). Therefore, it makes it easier to build anecdotal cases. We have extracted two special instances (*List 1* and *List 2*) each having 28 and 30 actors respectively(small enough to be easily interpreted and large enough to need computation!) which can be used to highlight conceptual strengths. *List 1* and *List 2* are enumerated in Table 3. The first instance consists of 28 actors from contemporary times. The second instance consists of 30 actors whose prime era was before 1970s. Since these two sets have fairly well known actors, we make our conceptual points using these instances.

Note that the analysis based on *alpha centrality* is useful only for connected interaction networks. When the graph is disconnected, the rankings across different components are not comparable. Therefore, we limit our instances to only connected interaction networks. So, we employ simple traversal techniques over the movie and actor lists to construct instances which are connected. To demonstrate the stability of the computational approach we construct interaction networks with 200, and 400 actors (11114 movies). As it is difficult to manually verify the efficacy, we employ the Kendall *tau* [10] distances (τ from now) to show how the method responds to the cases considered for the special instances mentioned before. As for the feasibility of our approach for very large graphs, it should be noted that the computation of an inverse is the most time consuming part; $O(n^3)$. Therefore, for most interaction networks like the completely cleaned, connected IMDB or DBLP, our approach can be used without any problem. For truly massive graphs like the web graph, exact computation like in this paper is not feasible. Developing a technique that can be used even on massive graphs is a challenging computational problem.

4.1 Experiments when outcomes are changed

The goal of these experiments is to evaluate the ability of alpha centrality (AC) and SVD to take into account outcomes while ranking the actors. We conduct our experiment as follows. Let Δ be

$R_1 = \text{AC}(\Delta, e)$	23 9 25 12 10 7 21 15 3 6 20 8 24 16 28 2 22 26 27 11 1 4 14 18 5 17 19 13
$R_2 = \text{SVD}(\Delta)$	9 23 25 12 7 10 21 15 3 20 8 6 24 28 16 2 22 26 27 11 4 1 14 5 17 18 19 13
$R_3 = \text{AC}(\Delta', e)$	25 28 23 9 12 10 7 21 15 3 16 20 6 8 24 2 22 26 27 11 14 1 4 18 17 5 19 13
$R_4 = \text{SVD}(\Delta')$	9 23 25 12 7 10 21 15 3 20 8 6 24 28 16 2 22 26 27 11 1 4 14 5 17 18 19 13

Table 4: Rankings of actors in List1 under different conditions

$\text{AC}(\Delta_{list1}, e)$	23 9 25 12 10 7 21 15 $3^{\clubsuit}$ 6 20 8 24 16 28 $2^{\diamondsuit}$ 22 26 27 11 1 4 14 18 5 17 19 13
$\text{AC}(\Delta_{list_2}, e)$	17 12 24 30 19 28 16 20 15 18 22 10 11 25 2 21 6 13 4 29 23 14 8 7 9 5 1 3 $27^{\clubsuit}$ $26^{\diamondsuit}$

Table 5: Rankings of Al Pacino$^{\diamondsuit}$ and Robert de Niro $^{\clubsuit}$ in both the lists

the matrix constructed according to the method in Section 3.2.1. Let e be the exogenous vector. We first evaluate the AC and SVD rankings of actors in a list (say List 1). The rankings are generated under four different conditions as follows:

- R_1 is the ranking by $\text{AC}(\Delta, e)$. Table 4 and Table 7 show the R_1 ranking for both the list of actors.
- R_2 is the ranking by $\text{SVD}(\Delta)$.
- R_3 is generated after making the following modifications to the original data. We pick two highly ranked actors in both the rankings, say A_1 and A_2. For each of the movies in which either of them appears, we artificially reduce the averaging rating by 2. We then pick the two middle-ranked actors, A_3 and A_4. We increase the ratings of those movies in which either of them appears by 2. Let Δ' be the corresponding matrix. Note that e does not change. Let R_3 be the ranking by $\text{AC}(\Delta', e)$.
- R_4 is the ranking obtained by $\text{SVD}(\Delta')$.

For the rest of the experiments we use R_1, R_2, R_3, R_4 to denote the rankings obtained by AC and SVD by this process for the different lists and different es considered in the experiments. The different rankings obtained for List 1 (See Table 4) are as in Table 4 (With $A_1 = 9, A_2 = 23, A_3 = 28, A_4 = 16$). The rankings are given in the ascending order of ranks; the actor with the first rank appears first and so on.

Let us now check how the two methods dealt with changes in outcomes (refer Table 4). One would expect the rankings of A_1 and A_2 to go down and those A_3 and A_4 to go up. Notice (comparing R_1 and R_3) that AC rankings of the both top actors (number 23 and 9) have gone down while those of the two chosen mid-ranked actors A3 and A4 (number 28 and 16) have gone up. Notice (comparing R_2 and R_4) that there is hardly a noticeable change in the SVD rankings before and after modification. This shows that the role played by the exogenous vector e in assigning the importance of outcomes is critical.

4.2 Experiments for relative importance across time periods

We did these set of experiments to see the variation in ranks of actors when movies (and actors) across different periods are considered. For e.g. some actors may be active say before 1970 so their ranks should be low when most of the movies (hyper-edges) in the interaction network are taken from beyond 1970. We noticed that the rankings were indeed dependant upon the periods from which movies were considered. For e.g. Robert De Nero and Al Pacino were ranked higher when movies till 2008 were considered, but were ranked lower when movies till 1980 were considered. The ranking results are shown in Table 5 where the first row shows the ranking for *list 1* and second row for *list 2*. These experiments were carried on List 1 (with all movies upto 2008) and List 2 (old actors and movies upto 1980).

4.3 Experimentation with e

As mentioned before, the role played by the exogenous vector is crucial. It is essentially used to capture the relative value created by different outcomes. Note that there are settings where the outcomes are easily seen to be either positive or negative. For example, in case of movies, a user

$R_1 = \text{AC}(\Delta, e)$	17 24 12 30 20 28 16 19 15 10 22 18 25 11 2 21 6 13 4 23 29 14 8 7 9 5 1 3 27 26
$R_3 = \text{AC}(\Delta', e)$	17 24 12 30 20 28 16 19 15 10 18 22 21 25 2 11 6 13 4 23 29 14 8 7 5 9 1 3 27 26

Table 6: Rankings of actors in List2 generated for Case 1

$R_1 = \text{AC}(\Delta, e)$	17 24 12 30 28 19 16 20 15 10 22 18 25 11 2 21 6 13 4 23 29 14 8 7 9 5 1 3 27 26
$R_3 = \text{AC}(\Delta', e)$	15 24 30 20 21 10 16 17 25 12 28 19 22 18 2 11 6 13 4 14 29 5 23 8 1 27 9 7 26 3

Table 7: Rankings generated of actors in List2 for Case 2

rating of 6 or below is usually an indication of its failure whereas a rating above 8 is usually an indication of its success. In this case, it is meaningful to associate positive and negative weights with different outcomes. However, in certain other settings, such as publications by researchers, every outcome in the form of publication in refereed conferences is essentially an outcome of some value. The value varies depending on forum and citations and so on. So, in this case, the weights associated with outcomes have to be positive and appropriately assign positive weights for different outcomes (for example, 100 citations is more than 10 times better than 10 citations!). However, ability to reflect the implied changes in rankings for R_1 and R_3 in these settings is different. We present our experience with these settings below.

Case 1. When all outcomes were positive. In this case, every movie is treated as having a non-zero positive outcome. Here, for a movie with rating R the outcome would be 2^R. The $e[i]$ for each actor i is uniformly set to 100. Let us consider the rankings R_1 and R_3 in this setting (for List 2) are shown in Table 6. The status of the actors 17 and 24 remains unchanged even though the average rating for their movies was reduced. However, even in this setting, the AC rankings reflect changes in the status of several actors like 21 and 11 do see the effect of change in outcomes.

Case 2. When the outcomes are mixed. This is the most general case where the outcomes are treated as positive and negative. For example, a movie whose average rating is 5 is deemed a failure and thus having a negative outcome, and a movie with a rating of 8 is considered to be a positive outcome. Specifically, we set the weights for different outcomes as follows. If a movie has the corresponding rating R greater than 7 then the associated was 2^R, else the outcome is -2^{8-R}. In this case, R_1 and R_3 were noticeably different. Here too, $e[i]$ for all actors is uniformly 100. Note that relative ranks of 17 (A_1) has dropped while that of 21 (A_4) has improved. The reason for the stability of 24 in the ranking is his close collaboration with actors 2 and 11 with whom he has acted in 14 movies. Table 7 are the ratings from experiment on List 2.

Case 3. When e vector is non-uniform for actors. We do this experiment to show how the exogenous vector can be leveraged to reflect sudden changes in the importance of individual actors, for example, due to an award or sudden raise in popularity etc. Notice that the actors with indices 27 (Robert De Nero) and 26 (Al Pacino) are appearing towards end in R_1. If we now increase the importance of these two actors and keep the importance of rest of the actors uniform, then, we notice that not only these two actors, but, also those who interacted with them (like Marlon Brando) benefit in the ranking. The new alpha-centrality ranking after changing importance of the two actors is denoted by R_5. Rankings are shown in Table 8.

4.4 Experiments on larger networks

As mentioned above, we performed experiments with lists of 200 and 400 actors. We use Kendell *tau* τ correlation [10] to see how rankings are correlated. Given two complete rankings, it computes a score in the range of $[-1, 1]$. If the score is close to 1, it means the two rankings are identical and if it is close to -1, it means the two rankings are near opposite orders. Now, in a large network, decreasing the value of outcomes of two top actors and increasing the outcome of two middle ranked actors should give rise to large shifts for almost all actors. Therefore, the two rankings are likely to look almost unrelated. This is because the rankings also reflect structure. So, the changes of the top two actors will affect a significant fraction of the actors. We notice that, for the AC rankings, the

$R_1 = \text{AC}(\Delta, e)$	17 24 12 30 28 19 16 20 15 10 22 18 25 11 2 21 6 13 4 23 29 14 8 7 9 5 1 3 27 26
$R_5 = \text{AC}(\Delta, e')$	27 26 24 25 20 17 28 30 16 15 10 12 22 2 19 14 29 18 6 4 11 21 13 1 8 23 7 9 5 3

Table 8: Rankings generated of actors in List2 for Case 3

Kendall *tau* measure was close to zero and for the SVD rankings they were in the range of 0.6. These experiments provide significant empirical evidence that our method accounts both the structure and the outcomes in its rankings; and, the SVD based approach is not as effective.

Case 1. Network of 200 actors. $\tau(R_1, R_3) = 0.007352$ and $\tau(R_2, R_4) = 0.68160$

Case 2. Network of 400 actors. $\tau(R_1, R_3) = 0.087$ and $\tau(R_2, R_4) = 0.6345$

5 Conclusions

In this work we presented an approach for node ranking which simultaneously leverages structural as well as value of the interactions. The main idea was to employ alpha centrality based ranking on the asymmetric adjacency matrix of a directed graph which has nodes corresponding to actors and special nodes corresponding to outcomes. Some of the interesting further work includes (i) use the methodology to develop a mechanism to rank agents in a service organization where interactions and outcomes (positive) are essential for success (ii) handle multi-dimensional outcomes by exploiting the exogenous vector innovatively, and (iii) explore richer set of algorithmic approaches. Empirically, it is essential to test this approach in other domains as well.

References

[1] Sitaram Asur, Srinivasan Parthasarathy, and Duygu Ucar. An event-based framework for characterizing the evolutionary behavior of interaction graphs. In *KDD*, pages 913–921, 2007.

[2] Philip Bonacich. Factoring and weighting approaches to clique identification. *Journal of Mathematical Sociology*, 2:113–120, 1972.

[3] Philip Bonacich and Paulette Lloyd. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23:191–201, 2001.

[4] Krishna Prasad Chitrapura and Srinivas R. Kashyap. Node ranking in labeled directed graphs. In *CIKM*, pages 597–606, 2004.

[5] Colin DeLong, Sandeep Mane, and Jaideep Srivastava. Concept-aware ranking: Teaching an old graph new moves. In *ICDM Workshops*, pages 80–88, 2006.

[6] Brenda Dietrich and Terry Harrison. Serving the services industry. *Operations Research and Management Science (OR/MS)*, 33(3), 2006.

[7] Taver H. Haveliwala. Topic-sensitive page rank. In *In the proceedings of the World Wide Web (WWW) Conference*, pages 517–526, 2002.

[8] J. E. Hirsch. An index to quantify an individual's scientific research output. *Proceedings of National Academy of Sciences*, 102(46):16569–16572, 2005.

[9] S Kameshwaran, Sameep Mehta, Vinayaka Pandit, Gyana Parija, Sudhanshu Singh, and N. Viswanadham. Analyses for service interaction networks with applications to service delivery.

[10] M. G. Kendall. Further contributions to the theory of paired comparisons. *Biometrics*, 11:43, 1955.

[11] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, 1999.

[12] Xiaoming Liu, Johan Bollen, Michael L. Nelson, and Herbert Van de Sompel. Co-authorship networks in the digital library research community. *Information Processing and Management*, 41:1462–1480, 2005.

[13] Mark Newman. Scientific collaboration networks: I. Network construction and fundamental results. *Physical Review*, 64, 2001.

[14] Mark Newman. Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality. *Physical Review*, 64, 2001.

[15] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web, 1998.

[16] T. H. Wei. *The algebraic foundations of ranking theory.* Cambridge University Press, London, 1952.

[17] Scott White and Padhraic Smyth. Algorithms for estimating the relative importance in networks. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 266–275, 2003.