Dependency Centrality from Bipartite Social Networks

Luke M. Gerdes
United States Military Academy
West Point, New York

Abstract
This paper introduces dependency centrality, a node-level measure of structural leadership in bipartite networks. The measure builds on Zhou et al.'s (2007) flow-based method to transform bipartite data and captures additional information from the second mode that existing measures of centrality typically exclude. Three previously published bipartite networks serve as test cases to demonstrate the extent of correlation among node-level centrality rankings derived from dependency centrality and those derived from canonical centrality measures: degree, closeness, betweenness, and eigenvector. Ultimately, dependency centrality appears to offer a novel means to measure importance in bipartite networks depicting social interactions.

Author
Luke M. Gerdes is an Assistant Professor in the Department of Behavioral Sciences & Leadership at the United States Military Academy in West Point, New York.

Notes
This work was supported by the Office of the Secretary of Defense, Minerva Initiative. The views expressed herein are those of the author and do not purport to represent the official policy or position of the United States Military Academy, the Department of the Army, the Department of Defense, or the United States Government.

Please send all correspondence to Luke M. Gerdes, Department of Behavioral Sciences & Leadership, United States Military Academy. Email: luke.gerdes@usma.edu

1. Introduction

Bipartite data has been prominent in social network analysis since the discipline's earliest days. Indeed, Freeman (2004) traces the origins of bipartite networks to the late 1890s, when John A. Hobson, a British newspaper man and economist whose work influenced Lenin and other prominent Marxists, gave birth to the much studied topic of interlocking corporate boards. He presented agent x organization data summarizing six men's co-membership on the boards of five prominent South African companies. In pre-WWII America, Davis et al (1941) continued the tradition of bipartite data and studied social ties among southern women by cataloguing their co-participation in luncheons and parties, leading to the creation of an agent x event dataset that remains widely studied to this day.

Despite its long history, bipartite data has always presented analytic challenges. Because standard measures of node-level centrality were designed to estimate actors' influence within a single mode (Freeman, 1979), they lack direct applicability to bipartite data. There have been several efforts to extend centrality to multi-modal data (Faust, 1997; Borgatti & Everett, 1997; Latapy et al, 2008), but these approaches have yet to gain wide acceptance. The majority of analysts continue to transform bipartite data into a single mode prior to its analysis, even though it is widely known that this approach can negatively impact the theoretical validity of measurements conducted in the resulting one-mode projection (Opsahl, 2013). For example, Wasserman and Faust (1994) note that degree centrality performs differently in one-mode mode data derived from transformation than in 'naturally occurring' one-mode data; projected data tends to contain a number of abnormally large cliques resulting from the inference of direct ties among all co-participants in relatively large events.

Other scholars have offered new approaches to data transformation that partially mitigate such pernicious effects (Padrón et al, 2011; Opsahl, 2013; Gerdes, 2014). Most notably, Newman (2001) designed an approach that divides the weight of ties formed through co-participation by the number of participants in each event, which causes the resulting one-mode network to value ties formed during large events less than those formed during small events. Thus, Newman's process restricts nodes' ability to acquire high degree centrality through participation in large cliques.

Although this and other similar transformation approaches are elegant, context can undermine their utility. More manpower is often required to solve hard problems or manage complex situations than is needed to deal with small problems and everyday situations. These circumstances suggest that a clique's output may be directly proportional to its size. Consequently, large cliques are not always indicative of fleeting relationships. Even when analysts implement more sophisticated methods to transform bipartite data to a single mode, there are often still grounds to question how well measurements of centrality conducted in the resulting one-mode projection describe node-level influence and importance.

In their effort to find a more nuanced means of comparing the similarity of actors in bipartite networks, Zhou et al (2007) point the way toward a new measure of centrality that captures additional information from the second mode. This approach mitigates some of the concerns regarding the validity of node-level centrality in projected data. The new measure, which this paper dubs "dependency centrality," is based on a derivation of structural equivalence among nodes and relies on a novel method that Zhou and his co-authors devised to transform bipartite networks into single mode networks. The first section following this introduction describes Zhou's method of projection, before the second section moves-on to discuss the calculation of dependency centrality. The third section situates this new measure within the context of existing scholarship by discussing dependency centrality's relationship to other measures, especially those described in Kleinberg's (1999) work on centrality among web pages. The fourth section presents results that compare dependency centrality to existing node-level heuristics, and the fifth and final section offers some brief conclusions about measurement's potential applications.

2. Transforming Bipartite Networks

Consider a bipartite network comprised of agents and events. By considering the agents as the holders of resources that flow through the network, Zhou et al determined that it is possible to infer an agent x agent matrix from these between-mode connections. Zhou's process has two stages: in the first, the resources flow from the agents to the events in direct proportion to each agent's degree. Thus, if an agent holds links to two events, half of his resources flow to each of these events; if an agent has links to three events, one third of his resources from to each of these events, and so on.

However, the resources do not remain parked on the events. Instead, they immediately flow back to the agents following the same redistribution rules. If an event has ties to four agents, then a quarter of the resources momentarily parked on the event flow to each of the four agents who share ties to the event. If an event has ties to and so on. Mathematics allows these two processes to be combined into a single step:

$$w_{ij} = \sum_{k:\, m_{ik};\, m_{jk} = 1} 1/(\deg_j * \deg_k),$$

where w_{ij} represents that strength of the tie that exists between i and j as the result of their co-participation in k events in the original bipartite matrix, M (Gerdes, 2014). Figure 1 also offers a graphic depiction of this flow-based process for a simple bipartite network containing four individuals (Tom, Dick, Bob, and Harry) and four events (A,B,C, & D).

It is worth noting that the flow metaphor explaining this data transformation process holds regardless of the unique features of the underlying bipartite network. Figure 2, which depicts a simple network containing an exceptionally skewed distribution of ties, offers a useful illustration. In the one-mode projection that results from the application of Zhou's process to this skewed network, every actor in the network receives an equal inbound tie to Tom, Dick, Harry, and Bob. However, because Ron participated in every event in the network, and because he was the only participant in events B through C, a smaller share of his resources flow to Tom, Dick, Harry, and Bob, while the majority of Ron's resources return to him. Thus, the redistributions inherent to Zhou's flow-based approach are not necessarily egalitarian. Instead they operate according to each network's unique underlying structure, which causes the redistributions in the network depicted in figure 2 to function akin to an exceptionally regressive tax: the resources held by the poorest agents (i.e. those with low degree in the two-mode network) are taken and redistributed evenly. The resources held by the single rich agent (i.e. the lone high-degree agent in the two-mode network) largely return to him, even as he receives an equal share of the resources the poorer agents held in the original bipartite network.

While the logic of Zhou's approach holds regardless of the specifics of the underlying two-mode network, astute observers will notice that this transformation algorithm only allows for binary data. Fortunately, this limitation can be overcome using the "pipes" approach that Newman (2004) utilized in his efforts to calculate second generation measures of centrality. The basic logic is simple: when nodes share an interaction of weight greater than 1, analysts should think of each additional unit of weight as an extra pipe between the nodes. If all pipes in the analysis are of equal diameter and are equally full, then the value of ties between ego and any alter can be expressed as a ratio of the total number of pipes that ego 'owns.' Thus, in figure 3, Tom sends a tie of 5/6 to Event A, because five pipes connect these two nodes, and Tom owns a total of 6 pipes. Similarly, in the second step of the transformation, Event A sends 4/9 of its resources to Harry, because four pipes connect these two nodes, and Event A 'owns' a total of 9 pipes. Using this convention, a generalization of Zhou's transformation process that allows for weighted data can be formalized as:

$$w_{ij} = \sum_{k:\, m_{ik};\, m_{jk} > 0} (m_{ik} * m_{jk})/(\deg_j * \deg_k),$$

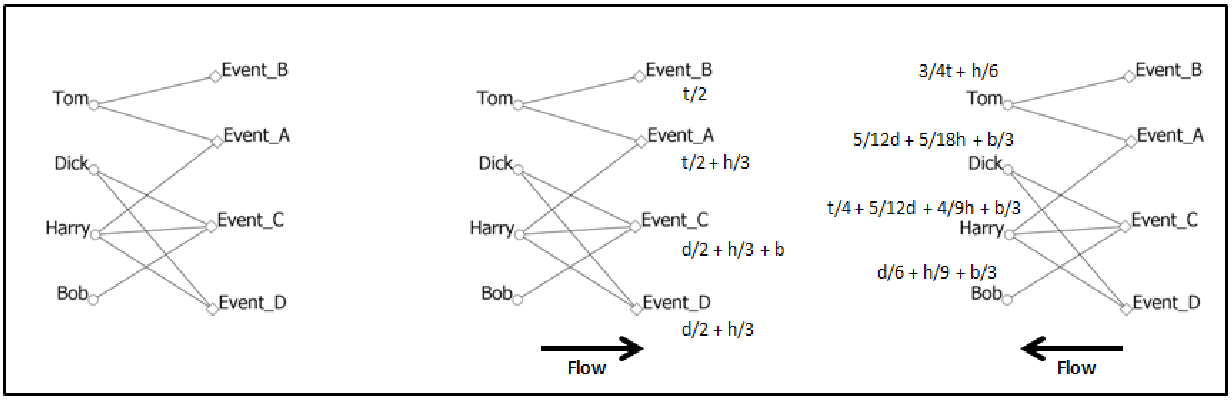
where w_{ij} is the strength of the relationship that agents i and j share as the result of co-participation in k events in the original two-mode matrix, M (Gerdes, 2014). Figure 3 offers a graphic depiction of the "pipes" generalization of Zhou's transformation process.

3. Toward Dependency Centralization

Even when generalized to accommodate weighted data, Zhou's process has limitations for analysts wishing to study traditional measures of centrality. Specifically, the method flattens the distribution of degree centrality by discounting the weight of new ties formed by high-degree agents. When applied to longitudinal data, the process would, consequently, undervalue the importance of novel partnerships, especially those forged to actors with high degree centrality. Moreover, the process produces directed graphs, which may not be appropriate in several circumstances involving agent *x* event data. For example, it is difficult to see how Tom could have interacted with Harry fewer times than Harry interacted with Tom through participation in the same event. It is worth noting that this criticism of directionality may become moot if the context of analysis dictates that agents hold different quantities of resources in the original bipartite graph, thereby enabling an exchange of "50 units" to be expressed as one-half of the resources held by an agent who started the transformation in possession of 100 units, and two-thirds of the resources held by an agent who started the transformation process in possession of 75 units.

Agent x Event

	Event A	Event B	Event C	Event D
Tom	1	1	0	0
Dick	0	0	1	1
Harry	1	0	1	1
Bob	0	0	1	0



Agent x Agent

	Tom	Dick	Harry	Bob
Tom	3/4	0	1/6	0
Dick	0	5/12	5/18	1/3
Harry	1/4	5/12	4/9	1/3
Bob	0	1/6	1/9	1/3

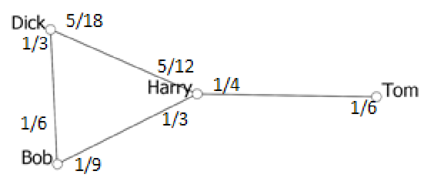
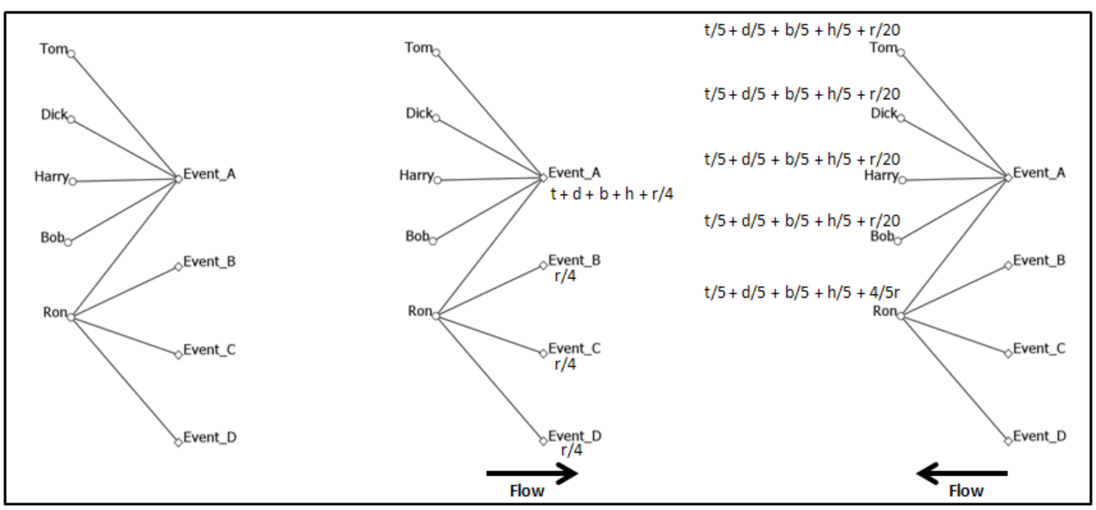


Figure 1: Zhou et al's Projection Process for Bipartite Data (Gerdes, 2014)

Agent x Event

	Event_A	Event_B	Event_C	Event_D
Tom	1	0	0	0
Dick	1	0	0	0
Harry	1	0	0	0
Bob	1	0	0	0
Ron	1	1	1	1



Agent x Agent

	Tom	Dick	Harry	Bob	Ron
Tom	1/5	1/5	1/5	1/5	1/20
Dick	1/5	1/5	1/5	1/5	1/20
Harry	1/5	1/5	1/5	1/5	1/20
Bob	1/5	1/5	1/5	1/5	1/20
Ron	1/5	1/5	1/5	1/5	4/5

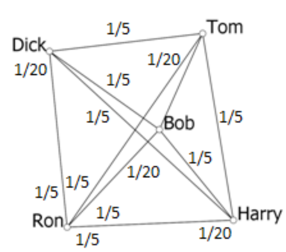
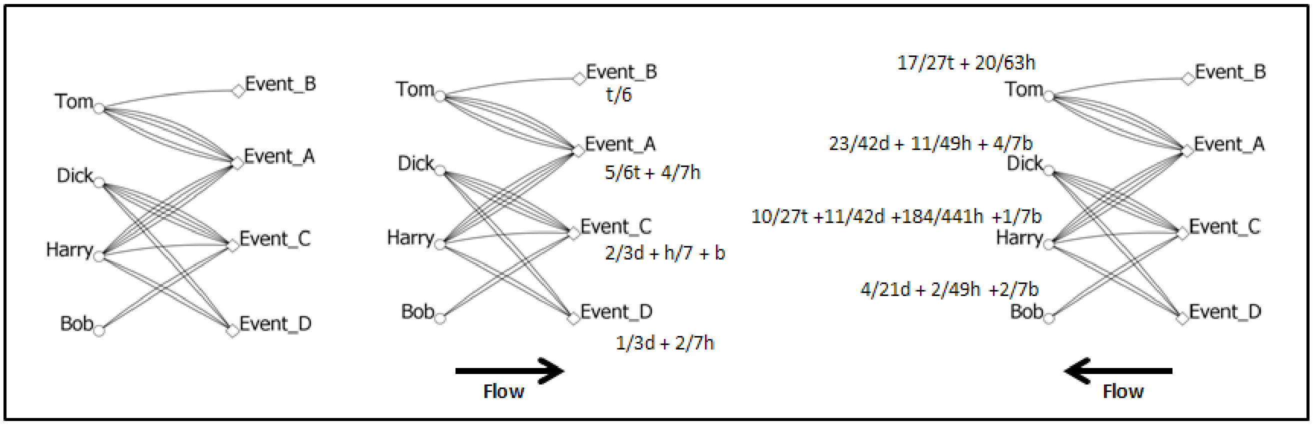


Figure 2: Zhou et al's Projection for Irregular Bipartite Data

Agent x Event

	Event A	Event B	Event C	Event D
Tom	5	1	0	0
Dick	0	0	4	2
Harry	4	0	1	2
Bob	0	0	2	0



Agent x Agent

	Tom	Dick	Harry	Bob
Tom	17/27	0	20/63	0
Dick	0	23/42	11/49	4/7
Harry	10/27	11/42	184/441	1/7
Bob	0	4/21	2/49	2/7

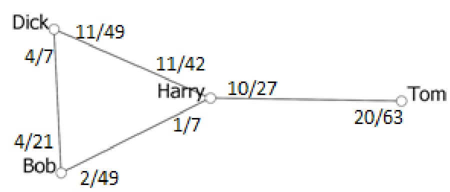


Figure 3: "Pipes" Generalization of Zhou et al's Projection Process for Bipartite Data (Gerdes, 2014).

Despite these limitations, Zhou's approach has one advantage unavailable from other methods of data transformation: this projection process enables analysts to calculate agent's structural dependence on one another. Structural dependence can be expressed as a ratio, varying between 0 (complete independence) and 1 (complete dependence), and which can be determined by dividing the number of ties an agent shares with an alter by the original agent's self-loop (i.e. the amount of resources that return to the agent at the end of the flow-based transformation). More formally:

$$D_{ij} = w_{ji}/w_{ii} \,,$$

where D_{ij} represents agent *i*'s dependence on agent *j* as the result of co-participation in events.[1]

Although Zhou et al's discussion of dependency ratios does not reference structural equivalence as introduced by Lorrain and White (1971), the two concepts are highly related. Nodes *i* and *j* are structurally equivalent if they possess identical in-ties and identical out-ties. In an agent *x* event network, node *i* is structurally dependent on node *j* if the later participates in *at least* the same events as the former. If *j* participates in additional events that do not involve *i*, the dependency of *i* on *j* remains unaffected, though *j* would then be only partially dependent on *i*. Thus, dependency ratios are effectively directed, non-reciprocal measurements of structural equivalence that can only be operationalized in bipartite networks that have been transformed using Zhou et al's flow-based transformation process to infer the uniquely weighted self-loops on which dependency ratios rely.

How are these ratios useful to analysts? According to Zhou et al, the dynamics of the professor-student relationship offer a useful illustration. If a professor co-authors a series of papers with several different graduate students, none of whom have previously published, each student will be entirely dependent on the professor (i.e. each student's *D* will reach the theoretical maximum of one), because they have no other publication partners.

[1]Zhou et al originally named this measure "independence," but as higher values indicate greater levels of dependence, their naming convention was counter-intuitive. The calculation remains unchanged from the original formulation, but has been re-titled to remain in-line with typical network conventions, as best demonstrated by the fact that greater indirect influence is typically conceptualized as an a higher measure of closeness centrality, rather than a lower measure of average path-length.

However, the professor will be independent of any single student (i.e. the professor's *D* will approach the theoretical minimum of zero) by virtue of his myriad publication partners. Thus, dependency ratios begin to capture notions of relative power.

Consequently, even in their raw state, these calculations are useful for assessing patterns across a host of applications. The student-mentor example highlights the applicability of dependency ratios to the study of citation networks, but the calculations can also help to improve the recommendations that firms provide to consumers based on the purchasing behavior of other similar customers (Zhou et al, 2007; Liu et al, 2009; Li et al, 2009). This approach may also offer a means to detect community structure in complex networks (Pan et al, 2010). Finally, as this paper contends, dependency calculations may serve to highlight leadership patterns within organizations, by assessing the extent to which supposed subordinates are structurally dependent on an organization's named leader.

Table 1: Pair-wise Dependencies

	T	D	H	B
T	0	0	0.333	0
D	0	0	1	0.400
H	0.375	0.625	0	0.25
B	0	1	1	0

In order to highlight these sorts of organizational dynamics, it is necessary to move from raw dependency calculations to a normalized node-level measurement of centrality. This task can be accomplished by conducting the sort of pair-wise comparison of dependency calculations listed in table 1, which shows all possible dependencies among participants in the one-mode network that appears in the lower left-hand corner of figure 1. This comparison ignores self-loops, since every actor is perfectly dependent on themselves.

Zhou et al propose that a node-level measurement of dependence can be determined by squaring the value of each cell, and then summing by column for each actor, such that overall dependence can be formalized as:

$$\sum_j (w_{ji}/w_{ii})^2 .$$

However, this non-linear measure is somewhat problematic. Given that it is impossible to generate a negative dependency ratio, it remains unclear why it is appropriate to square each measurement. Since cell values range between 0 and 1, this action seems to artificially inflate differences between actors. Values at the theoretical minimum (i.e. zero) and the theoretical maximum (i.e. one) remain unaffected by the squaring process, while decimalized values between these extremes become smaller when multiplied against themselves. For example, Tom's raw comparison to Harry, initially valued at one-third, becomes one-ninth when squared, while Dick and Bob's comparison to Harry, which are both valued at one, retain their original value despite squaring. Simply put, squaring can distort results for partially-dependent actors, by making such individuals appear less bound to structurally independent 'mentors' than raw dependency calculations would suggest.

The node-level measurement proposed by Zhou et al is also problematic because it lacks comparability across networks. This deficiency can be illustrated by amending the student-mentor example used to demonstrate the interpretation of dependency calculations. Assume that the goal of a study is to measure the annual change in a professor's dependence on students. Unless the professor under analysis publishes with the same number of students each year, his maximum possible dependency score will fluctuate based on the number of yearly co-authors. If he authors with five students in the first year of the study, the professor's maximum node-level score will be five, but if he authors with four students in the study's second year, his maximum node-level score will drop to four. Therefore, if the professor held a composite score of 0.5 in both years, it would be incorrect to conclude that his dependence on students was stable from year-to-year, because this score represents 2.5 percent more of the theoretical maximum in the second year ((0.5/5) – (0.5/4) = 0.025). The node-level measurement proposed by Zhou et al is ultimately of limited utility because it fails to take the size of the network into account.

Fortunately, a more intuitive, normalized node-level measurement can be calculated from the sort of pair-wise comparisons of dependency ratios listed in table 1. Begin by taking the column sums of the pair-wise dependency table. This process yields a raw cumulative measurement of the extent to which all other nodes are dependent on a given actor. For example, Harry holds a raw cumulative value of 2.33, because Tom is 1/3 dependent on Harry, while Dick and Bob are both entirely dependent on Harry. Next, divide this column sum by *n-1* in order to determine the mean extent to which all other nodes in the one-mode network are dependent on a given actor. Harry's score now becomes approximately 0.78. Freeman (1979) implemented the standard node-level normalization process division by *n-1* to ensure that centrality scores derived from binary data range between the added benefit of cross-comparability among networks of different sizes. This "dependency centrality," which is effectively an implementation of in-degree centrality on matrix that has been processed via Zhou's flow-based transformation algorithm, is computed by:

$$D_C = \frac{\sum_{j \neq i} \frac{w_{ji}}{w_{ii}}}{(n-1)},$$

where D_c represents dependency centrality; w_{ji} represents the weight of a tie from actor *j* to actor *i*; w_{ii} represents the weight of actor *i*'s self-loop, and *n* represents the total number of agents in the network.

When viewed in the context of bipartite agent *x* event data, this measurement seems to capture structural leadership, since agents who score highly in dependency centrality have a large number of followers who participate in events only alongside the central individual. Conversely, individuals with low, but non-zero, dependency centrality scores appear to be "followers," who typically participate in events alongside other more active individuals. Agents who are non-isolates in the original two-mode network, but who display dependency scores of zero appear to be lone wolves, who operate absent both social leaders and social followers.

These interpretations suggest that dependency centrality offers a useful means of identifying key actors from bipartite data, which is among the most common classes of network data. However, as the following section on the relationship between this measure and existing scholarship makes clear, dependency centrality is best suited for bipartite networks that depict some aspect of social behavior. Other measures appear better suited to study bipartite networks that exist outside the human environment.

4. Relationship to Related Work

In addition to dependency centrality's aforementioned relationship to structural equivalence, the new measure brings to mind Kleinberg's (1999) study of the role network structure plays in the search for authoritative web pages—a work that dismisses in-degree as a valid means for finding authoritative nodes. Specifically, Kleinberg argues that in-degree can inflate the importance of superfluous web pages because many hyperlinks, such as those that allow users to navigate back to an organization's main page, are not topically-oriented and thus have nothing to do with the interrelationships among pages containing similar content. Moreover, when "authority" is defined among a sub-set of pages that contain a specific user-determined search-term, bad tagging can further problematize in-degree's ability to capture the concept, because the top returns of searches ordered by in-degree often fail to match the user's intent. Kleinberg illustrates this point by noting that Amazon.com featured among pages with the largest number of in-links for pages that contained 'information' on the "Java" computer language. This result occurred because the breadth of Amazon's retail catalogue makes the site "universally popular," causing it to "have large in-degrees regardless of the underlying query topic" (p. 610-611). Such considerations led Kleinberg to conclude that eigenvector-based heuristics, which attribute importance to nodes based on the structural importance of their alters and not merely these alters' number, offer better means to locate authoritative web pages than the naïve results obtains from in-degree.

Although Kleinberg is concerned with one-mode networks of web pages, as opposed to bipartite networks, there is a temptation to assume that his conclusions about the relative merits of in-degree and eigenvector centralities are generalizable to all networks, regardless of their underlying content and structure. But context matters. Bipartite social networks appear to be largely devoid of the characteristics that led Kleinberg to dismiss in-degree as a valid measure of node-level importance.

Consider an agent *x* event network depicting social functions. Provided that the analyst has given some thought to issues of data collection and is not attempting to infer centrality based on co-attendance of the Super Bowl, stadium concerts, or other events that boast large crowds, it is difficult to think of an example in which a person's role in the social network is divorced from the extent of their participation in events along the same lines that caused Amazon.com to appear falsely central when Kleinberg used in-degree to rank web pages associated with the search-term "Java." People have limited capacity to attend social events, and thus do not typically form false associations with events in the same manner or on the same scale as monolithic websites that boast virtually limitless capacity for tie formation, allowing them to hold indiscriminate associations with topics of every variety. If analysts follow sound data collection protocols in building a bipartite social network, then they can have confidence that any agent who participates in every event in the network plays an important social role.

Similarly, properly designed data collection protocols will bypass information that is akin to the navigational shortcuts (e.g. "return to home" hyperlinks) that problematize efforts to use in-degree to rank web pages based on the structure of their ties to other online portals. Returning to the example of an agent x event network depicting social functions, an analyst would only need to worry about the social equivalent of non-thematic navigational ties if the data neglected to distinguish between social functions and other types of gatherings, such as corporate meetings. The previous examples make a simple but important point: different analytic tools are appropriate for different analytic contexts. Just because in-degree functions poorly for online networks does not mean that this type of centrality functions poorly for social networks.

It is also worth examining the other side of the coin by briefly discussing the appropriateness of implementing on dependency matrices the sort of eigenvector-based heuristics that Kleinberg recommends to rank the results of online search queries. Although some prominent network theorists have recommended against the use of eigenvector-based centralities on directed data (Bonacich 1972, Valente et al, 2008), such measurements can be conducted on any square matrix, and a number of approaches have been developed to examine in-links and out-links (see Bjelland et al, 2010, for a review). Thus, mathematics does not preclude the calculation of various eigenvector-based centralities on the inherently non-symmetric dependency matrices.

However, directed versions of eigenvector typically involve transforming the adjacency matrix under study into a stochastic matrix, which transposes the position of the rows and columns, before transforming cell values through division by individual column totals (Page et al, 1998). Given that the dependency matrix is already a derivative of the derivative of the original bipartite data, it is not entirely clear what sort of information an eigenvector-based measurement of centrality would capture when applied to the dependency matrix. The resulting measurements may simply be too far removed from the original data to accurately characterize agents' influence and importance.

A more intuitive result could be obtained by transforming the original bipartite matrix using a process other than Zhou's, and then implementing an eigenvector-based measure of centrality in that matrix. The following section on testing implements this approach to eigenvector-based heuristics in its effort to compare the results of dependency centrality to several other existing measures of node-level importance.

5. Evaluating Dependency Centrality

The nature of dependency centrality complicates comparative evaluations. Because the new measure relies on the positive self-loops derived from the projection process innovated by Zhou et al, this measurement explicitly considers information that originates in a second mode. Therefore, it is intended to be applied only to bipartite networks. Consequently, dependency centrality lacks the near-universal generalizability of traditional measures of centrality (e.g. degree, closeness, and betweenness centralities), which were originally intended to measure node-level influence within a single mode (Freeman, 1979). Simply put, dependency centrality is conceptually distinct from the canonical measures of centrality, and the new measure is difficult to compare to these centralities because they attempt to measure different things.

This theoretical distinction has practical consequences for evaluation. As discussed in a preceding section, it is difficult to have full confidence in any standard measures of centrality derived from the one-mode networks resulting from the flow-based projection process of Zhou et al. Efforts to gauge the relative performance of dependency centralization must, therefore, rely on two distinct transformation processes: one to calculate standard network measures, and another that is only useful for calculating dependency centrality. Stated differently, it is necessary to derive benchmarks for comparison by processing a bipartite network using a conventional approach to data transformation, before deriving dependency centrality scores by re-processing the same bipartite network using the flow-based projection process of Zhou et al.

This paper implements a two-pronged approach to testing in regards to three previously published data sets. These are drawn from various subject matter in an effort to situate dependency centrality in a broad intellectual context accessible to researchers from diverse disciplines. First, the paper evaluates the data that Davis et al (1941) collected on the participation of 18 southern women in 14 social events (see also, Breiger, 1974). Figure 4 offers a visualization of this network (generated via Csardi & Nepusz, 2006). Next, the paper evaluates the performance of dependency centrality within the context of dark networks by testing the measure using agent x event data that Center for Computational Analysis of Social and Organizational Systems, a research group at Carnegie Mellon University, collected on the participation of 18 al-Qaida members in 25 functional tasks underlying the 1998 bombings of the U.S. Embassies in Nairobi, Kenya, and Dar es Salaam, Tanzania (CASOS, 2008; Gerdes, 2008). Figure 5 offers a visualization of this network (generated via Csardi & Nepusz, 2006). Since the first two sets of test data are binary and relatively small, the final evaluation tests the performance of dependency centrality within the context of a larger, weighted bipartite network. Data that Opsahl (2013) collected on "Facebook-like forums" utilized by students at University of California at Irvine serves this purpose. In this dataset, 899 students are tied to 522 topical forums, and tie weights are determined by the number of times that users posted to each forum. Figure 6 offers a visualization of this network (generated via Csardi & Nepusz, 2006). In all three visualizations, white circles represent agents, and gray squares represent agents.

With the selection of test cases complete, data processing was accomplished using original scripts written in the R Language and Environment for Statistical Computing (R Core Team, 2012). In addition to the flow-based transformation process that was applied to all three of the bipartite networks used in testing, the southern women and al-Qaida datasets were 'folded' into symmetrical agent x agent networks, by multiplying the original bipartite networks against their transpose. Since the application of matrix multiplication to weighted bipartite networks produces one-mode data containing badly distorted tie-strengths (Padrón et al, 2011), a "one-way sum" projection was applied to the data on "Facebook-like forums" in order to produce a weighted agent x agent network for testing.

As its name suggests, the one-way sums approach to projection works via addition. The strength of agent x agent ties is determined by conducting a pair-wise comparison of agents' weighted two-mode interactions. Specifically, this projection process assumes that an agent "sends" ties to all of the other agents who co-participate in the same event as the sender; if agent A held a tie of strength 1 to an event, and agent B held a tie of strength 3 to the same event, then the tie from A to B would be valued at 1, and the tie from B to A would be valued at 3. When individuals co-participate in multiple events, the total weight of the agent x agent relationship is found by summing across events. Thus, the one-way sums projection that was applied to the forums data can be formalized as:

$$w_{ij} = \sum_{k: m_{jk} > 0} m_{ik} ,$$

where w_{ij} is the weight between node *i* and node *j* in the resulting agent *x* agent matrix, and *k* represents the event(s) in which *i* and *j* co-participated in the two-mode matrix, *M* (Opsahl, 2009; Padrón et al, 2011; Gerdes, 2014).

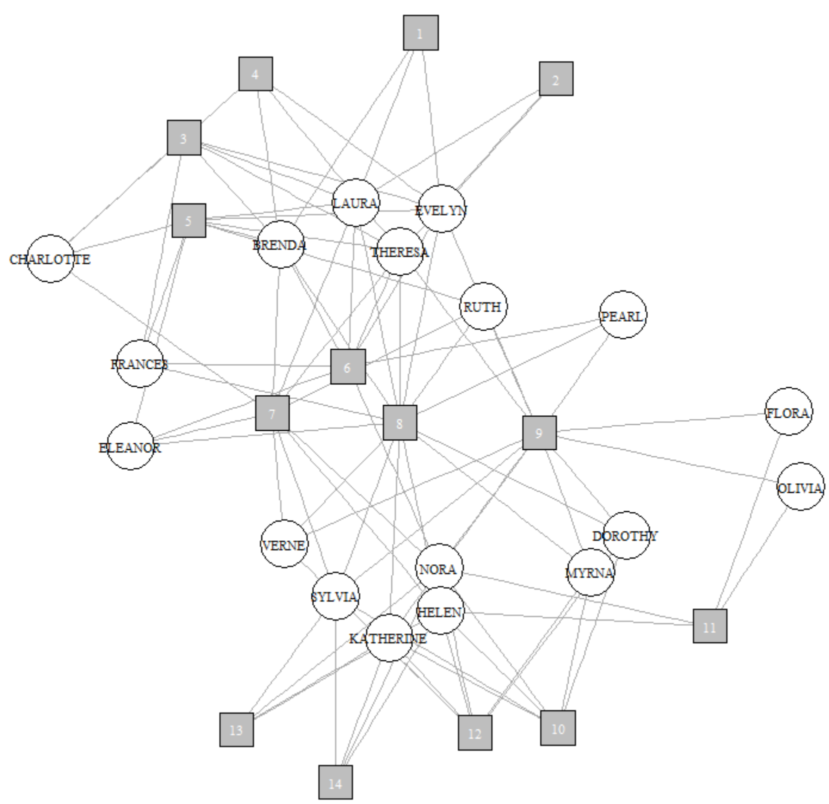


Figure 4: 18 Southern Women's Participation in 14 Events

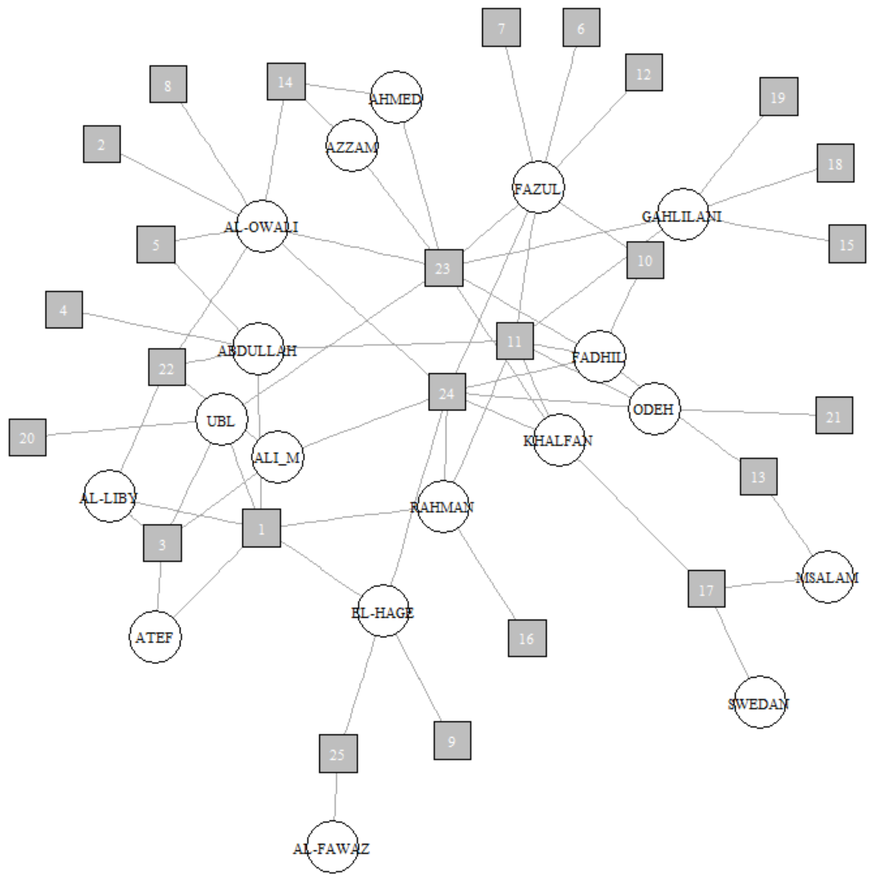


Figure 5: 18 al-Qaida Members' Participation in 25 Events

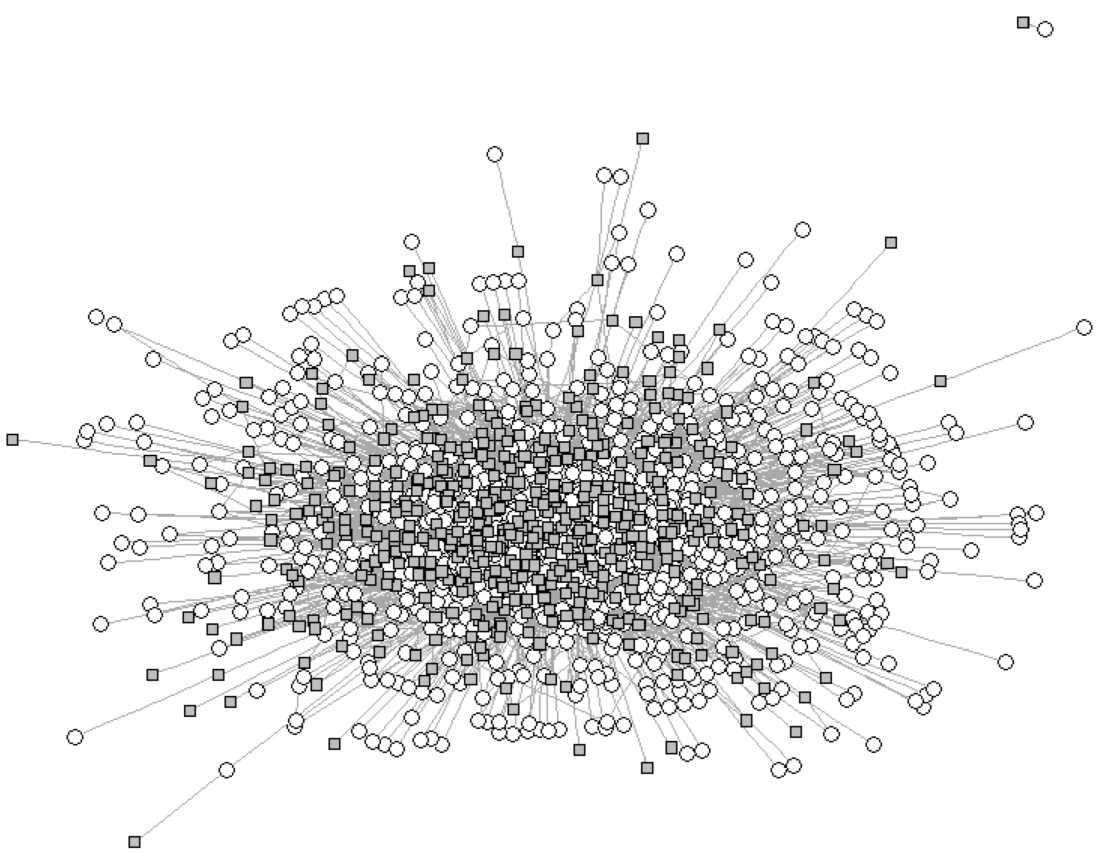


Figure 6: 899 Students' Participation in 522 Facebook-like Forums

Beyond data processing, this paper also utilized the R language to measure node-level centrality. Csardi and Nepusz's igraph package (2006) provided normalized measurements of degree, closeness, and betweenness in the one-mode networks derived from folding and one-way sums. Igraph also provided eigenvector-based measures for the networks derived via conventional approaches to transformation. Bonacich's eigenvector centrality, which was designed for symmetric matrixes, was applied to the networks derived from folding. However, because the one-way sums process is inherently directional, a fairer comparison of dependency centrality's relative performance could be obtained by applying an inherently directional eigenvector-based heuristic to the one-way sums transformation of the "Facebook-like" data. Page et al's (1999) algorithm, which underlies Google's approach to ranking the returns of online searches based on web pages' inbound links, serves that purpose. Finally, an original script was applied to the one-mode networks derived from Zhou et al's flow-based transformation process in order to measure dependency centrality.

Tables 2, 3, and 4 respectively summarize results for the southern women data, the al-Qaida data, and the forums data. These tables present both the raw scores and node rankings, and results are ordered by nodes' rank in dependency centrality. Given the size of the forums data, table 4 presents information on only the 20 individuals who rank highest in dependency centrality.

The results presented in tables 2 through 4 demonstrate that dependency centrality identifies different top-ranking agents than the canonical measurements of centrality. However, these differences tend to flatten-out for low-ranking individuals, especially when comparing dependency centrality with degree centrality or eigenvector-based centralities. Indeed, these three measures identically ranked the bottom 5 of the 18 southern women studied by Davis et al. Thus, dependency centrality performs differently than existing measures in the selection of "key" individuals and provides a unique measurement of node-level importance that is not captured by existing centrality measures.

However, it is also useful to consider the larger picture and assess the overall extent of correlation between the new measure and existing conceptions of centrality. A whole-network measure of centrality correlation is best accomplished by comparing agents' rank, since it is unsurprising that different measurements produce different raw values. Therefore, this paper relied on Spearman's Rho, which is akin to a Pearson correlation for ranks (McDonald, 2009a). Table 5 summarizes the extent of rank-correlation among dependency, degree, closeness, betweenness, and eigenvector centralities across all three of the test networks. Critical values of Rho were determined by following the widely accepted practice of approximating true values by using Student's t distribution with $df = N - 2$ (Ramsey, 1989).

Table 2: Comparing Measures of Centrality for Southern Women

Node ID	Depend.	Rank (Dep.)	Degree	Rank (Deg.)	Close.	Rank (Cls.)	Between.	Rank (Btw.)	Eigen.	Rank (Eigen.)
NORA	0.6026	1	5.5294	4	1.0000	4	0.0239	3	0.8099	6
THERESA	0.5763	2	6.7059	1	1.0000	4	0.0421	1	1.0000	1
EVELYN	0.5395	3	5.8824	2.5	1.0000	4	0.0360	2	0.8961	2
SYLVIA	0.5027	4	5.8824	2.5	1.0000	4	0.0150	5	0.8717	3
HELEN	0.4999	5	4.9412	7	1.0000	4	0.0138	6	0.7478	7
BRENDA	0.4637	6	5.4118	5	0.8947	13.5	0.0062	9	0.8518	4
LAURA	0.4510	7	5.2941	6	0.8947	13.5	0.0047	10	0.8415	5
KATHERINE	0.4203	8	4.8235	8	0.9444	9.5	0.0046	12	0.7268	8
VERNE	0.3224	9	4.5882	10	1.0000	4	0.0134	7	0.6920	10
RUTH	0.3218	10	4.7059	9	1.0000	4	0.0217	4	0.7138	9
MYRNA	0.3167	11.5	4.1176	12.5	0.9444	9.5	0.0046	12	0.6189	12.5
DOROTHY	0.3167	11.5	4.1176	12.5	0.9444	9.5	0.0046	12	0.6189	12.5
ELEANOR	0.2951	13	4.2353	11	0.8947	13.5	0.0031	14.5	0.6754	11
FRANCES	0.2787	14	3.7647	14	0.8947	13.5	0.0031	14.5	0.6087	14
PEARL	0.2402	15	3.6471	15	0.9444	9.5	0.0092	8	0.5595	15
CHARLOTTE	0.2329	16	2.8235	16	0.7391	18	0.0000	17	0.4810	16
OLIVIA	0.1659	17.5	1.6471	17.5	0.7727	16.5	0.0000	17	0.2322	17.5
FLORA	0.1659	17.5	1.6471	17.5	0.7727	16.5	0.0000	17	0.2322	17.5

Table 3: Comparing Measures of Centrality for al-Qaida Members (18 Agents)

Node ID	Depend.	Rank (Dep.)	Degree	Rank (Deg.)	Close.	Rank (Cls.)	Between.	Rank (Btw.)	Eigen.	Rank (Eigen.)
AL-OWALI	0.2531	1	2.3529	4	0.8095	3	0.0982	5	0.8473	4
KHALFAN	0.1943	2	2.5882	1.5	0.8500	1	0.3576	1	0.9513	3
UBL	0.1900	3	1.7647	6.5	0.8095	3	0.1171	4	0.5711	9
FADHIL	0.1701	4	2.5882	1.5	0.8095	3	0.1428	3	1.0000	1
FAZUL	0.1590	5	2.4706	3	0.7727	6.5	0.0388	8	0.9937	2
EL-HAGE	0.1572	6	1.5294	9.5	0.7727	6.5	0.2599	2	0.5334	11
ABDULLAH	0.1505	7	1.7647	6.5	0.7727	6.5	0.0468	7	0.6420	7
AL-LIBY	0.1504	8	1.2941	12	0.5862	14	0.0029	13	0.3923	14
ALI_M	0.1350	9	1.5294	9.5	0.7391	9	0.0346	9	0.5571	10
RAHMAN	0.1290	10	2.1176	5	0.7727	6.5	0.0599	6	0.8034	5
MSALAM	0.1069	11	0.3529	16	0.5000	16	0.0037	11.5	0.1260	16
ATEF	0.0933	12	0.9412	15	0.5667	15	0.0000	16	0.2787	15
GHAILILANI	0.0927	13	1.5294	9.5	0.6800	10.5	0.0106	10	0.6406	8
AHMED	0.0912	14.5	1.0588	13.5	0.6071	12.5	0.0000	16	0.4175	12.5
AZZAM	0.0912	14.5	1.0588	13.5	0.6071	12.5	0.0000	16	0.4175	12.5
ODEH	0.0728	16	1.5294	9.5	0.6800	10.5	0.0037	11.5	0.6692	6
SWEDAN	0.0505	17	0.2353	17	0.4857	17	0.0000	16	0.0673	17
AL-FAWAZ	0.0164	18	0.1176	18	0.4474	18	0.0000	16	0.0333	18

Table 4: Comparing Measures of Centrality for Users of a Facebook-like Forum (20 of 899 Agents)

Node ID	Depend.	Rank (Dep.)	Degree	Rank (Deg.)	Close.	Rank (Cls.)	Between.	Rank (Btw.)	Page	Rank (Page)
100	2.4341	1	91.3619	1	0.3037	1	0.1751	1	0.0098	1
18	2.2034	2	39.6080	2	0.2912	9	0.0535	4	0.0050	9
275	1.0471	3	29.4410	5	0.2837	29.5	0.0187	14	0.0032	44
67	0.9920	4	31.4666	4	0.3007	2	0.0948	2	0.0075	2
102	0.8776	5	17.7227	17	0.2719	148	0.0068	56	0.0019	159
47	0.8764	6	25.8797	6	0.2903	10	0.0324	7	0.0053	8
626	0.8446	7	22.2082	11	0.2900	11	0.0292	9	0.0044	13
17	0.8273	8	23.7528	7	0.2803	53	0.0150	20	0.0033	38
325	0.8110	9	19.4076	14	0.2773	77.5	0.0131	24	0.0027	64
290	0.8059	10	23.6125	8	0.3002	3	0.0805	3	0.0064	4
810	0.7255	11	35.1726	3	0.2942	7	0.0452	5	0.0056	6
287	0.6853	12	22.6604	10	0.2821	38.5	0.0107	30	0.0036	27
358	0.6523	13	23.5412	9	0.2807	47.5	0.0119	27	0.0034	36
173	0.5837	14	13.3641	31	0.2741	116.5	0.0109	28	0.0025	80
195	0.5680	15	21.6292	12	0.2823	37	0.0172	15	0.0032	45
172	0.5669	16	14.6314	24	0.2845	27	0.0190	13	0.0038	21
582	0.5633	17	11.1102	39	0.2795	57	0.0206	11	0.0031	48
625	0.5179	18	17.1481	19	0.2803	53	0.0171	16	0.0031	49
206	0.5046	19	17.7684	16	0.2866	19	0.0202	12	0.0046	11
16	0.4717	20	12.5690	35	0.2785	65	0.0035	108	0.0025	75

Bonferroni's correction for multiple comparisons was also applied (McDonald, 2009b) in order to determine that each of the upper three 4 x 4 tables listed below is significant at an alpha of 0.05.

The results of the correlation analysis largely confirm what previous studies have already shown: measures of centrality are often highly correlated (Coleman et al 1966; Burt, 1987; Bolland, 1988; Faust, 1997; Valente & Forman, 1998; Lee, 2006; Valente et al, 2008). Dependency centrality is, by mean across the three test networks, most closely correlated with degree centrality, at 90.5 percent. Betweenness performs similarly and is 87.9 percent correlated with dependency centrality, while closeness is only 83.8 percent correlated with dependency centrality. When Page's approach is treated as comparable with Bonacich's approach, eigenvector-based measures of centrality are only 84.5 percent correlated with dependency centrality.

It is worth noting that these patterns did not hold across the three individual test cases that contributed to this average. In the al-Qaida data, degree was less correlated with dependency centrality than either closeness or betweenness, and eigenvector centrality performed aberrantly, scoring only a 67.1 percent correlation with dependency centrality, even though comparable correlations were larger than 90 percent in the other two test networks. Although these results dictate that it is difficult to determine exactly how closely dependency centrality mimics the results of other measures of centrality, it is clear that the new measure performs differently. This finding holds both in terms of the selection of key actors as well as for correlations of node-level importance that consider all agents in the network. Ultimately, dependency centrality provides a measurement of structural leadership that is not fully captured by existing measures.

Table 5: Correlations of Node-Level Rank as Determined by Dependency, Degree, In-Degree Closeness, Betweenness, and Eigenvector Centralities

	Southern Women N = 18				
	Depend.	Degree	Close.	Between.	Eigen.
Depend.	1	0.97258	0.76232	0.85937	0.95036
Degree		1	0.70454	0.84009	0.99327
Close.			1	0.90882	0.66326
Between.				1	0.81578
Eigen.					1
	al-Qaida Members N = 18				
	Depend.	Degree	Close.	Between.	Eigen.
Depend.	1	0.81706	0.85895	0.85391	0.67149
Degree		1	0.95024	0.85663	0.96572
Close.			1	0.91478	0.87978
Between.				1	0.75056
Eigen.					1
	Facebook-like Forums N = 899				
	Depend.	Degree	Close.	Between.	Eigen.
Depend.	1	0.92533	0.89331	0.92518	0.91228
Degree		1	0.96826	0.90876	0.97648
Close.			1	0.90624	0.98598
Between.				1	0.92203
Eigen.					1
	Mean Across Test Networks				
	Depend.	Degree	Close.	Between.	Eigen.*
Depend.	1	0.90499	0.83820	0.87949	0.84471
Degree		1	0.87435	0.86849	0.97849
Close.			1	0.90995	0.84301
Between.				1	0.82946
Eigen.					1

** This mean treats the undirected eigenvector centrality scores and the directed Page centrality scores as comparable.*

6. Conclusions

This paper summarized a flow-based method to transform bipartite data into one-mode networks that was previously described by Zhou et al. This paper also presented a generalization that allows this method to be applied to weighted networks. However, this paper's primary contribution is to present dependency centrality, a new measure that the existing literature on centrality does not appear to describe.

This measure performed differently than widely-implemented measures of centrality across three test networks. Thus, dependency centrality appears to estimate aspects of structural leadership that are overlooked by existing means of determining node-level importance.

As the measure's name implies, dependency centrality more accurately scores the extent to which nodes in a bipartite network are structurally reliant on high-ranking nodes. For agent *x* event networks, high dependency centrality scores suggest that an agent has a large share of individuals who participate in events only alongside the central actor, and low dependency centrality scores that are greater than zero suggest that an agent tends not to participate in events absent more central individual(s). In the unique case that an agent is not an isolate in the original bipartite network, but has a dependency centrality of zero, it indicates that the agent is a lone wolf, who neither follows a central leader nor possesses adherents of his own. Thus, dependency centrality appears to capture the dynamics of mentormentee, recommender-recommendee, and other similar relationships.

However, leadership and importance are diffuse concepts, the definitions of which often change based on context, and diverse measures are necessary to assess different aspects of these concepts in different analytic environments. Dependency centrality is a reasonably specialized measurement that only functions in bipartite networks and cannot be applied in 'natural' one-mode networks. Moreover, Kleinberg's discussion of the features of online networks, which are replete with dubious connections and contain nodes that have limitless capacity to acquire alters, strongly suggests that dependency centrality should only be implemented in social environments and other contexts in which nodes in the bipartite graph have limited ability to form connections. While dependency centrality ultimately lacks the broader generalizability of degree, closeness, betweenness, and eigenvector-based measures of centrality, the new measure provides a distinct and valid means to assess node-level importance that estimates the extent to which nodes are structurally reliant on one another. Dependency centrality has clear utility for analysts attempting to determine organizational structure, analysts studying recommendation patterns among consumers, and analysts examining other socially-oriented issues that feature a leader-follower dynamic.

References

Bjelland, J., Burgess, M., Canright, G., & Engø-Monsen, K. (2010). Eigenvectors of directed graphs and importance scores: dominance, T-Rank, and sink remedies, *Data Mining and Knowledge Discovery*, Vol. 20, pp. 98-151.

Bolland, J.M. (1988). Sorting out centrality: An analysis of the performance of four centrality models in real and simulated networks. *Social Networks*, Vol. 10, No. 3, pp. 233-253.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Mathematical Sociology*, Vol. 2, No. 1, pp. 113-120.

Borgatti, S.P., & Everett, M.G. (1997). Network analysis of 2-mode data. *Social Networks*, Vol. 19, pp. 243-269.

Breiger, R. (1974). The duality of persons and groups. *Social Forces*, Vol. 53, pp. 181-190.

Burt, R.S. (1987). Social contagion and innovation: Cohesion versus structural equivalence. *American Journal of Sociology*, Vol. 92, pp. 1287-1335.

Center for Computational Analysis of Social and Organizational Systems (CASOS). (2008). Tanzania-Kenya-imoon.xml. Data available online: http://www.casos.cs.cmu.edu/computational_tools/datasets/internal/tanzania_kenya/index11.php.

Coleman, J.S., Katz, E., & Menzel, H. (1966). *Medical Innovation: A Diffusion Study*. New York, NY: Bobbs Merrill.

Csardi, G. & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, Vol. 1695.

Davis, A., Gardner, B.B., & Gardner, M.R. (1941). Deep South: A Social Anthropological Study of Caste and Class.. Chicago, IL: University of Chicago Press. Data available online: http://members.ozemail.com.au/~dekker/bipartite/.

Faust, K. (1997). Centrality in affiliation networks. *Social Networks*, Vol. 19, pp. 157-191.

Freeman, L.C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, Vol. 1, pp. 215-239.

Freeman, L.C. (2004). *The Development of Social Network Analysis: A Study in the Sociology of Science*. Vancouver, BC: Empirical Press, pp. 18-19.

Gerdes, L.M. (2008). Codebook for network data on individuals involved with terrorism and counterterrorism. *Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report CMU-ISR-08-136*. Available online: http://www.casos.cs.cmu.edu/publications/papers.php.

Gerdes, L.M. (2014). MAPPing dark networks: A data transformation method to study clandestine organizations. *Network Science*, Vol. 2, No. 2, pp. 213-253.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, Vol. 46, No. 5, pp. 604-632.

Latapy, M., Magnien, C., & Del Vecchio, N., (2008). Basic notations for the analysis of large two-mode networks. *Social Networks*, Vol. 30, No. 1, pp. 31-48.

Lee, C. (2006). Correlations among centrality measures in complex networks. *Physics and Society*.

Li, M., Dias, B., Jarman, I., El-Deredy, W., & Lisboa, P. (2009). Grocery shopping recommendations based on basket-sensitive random walk. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY:

Association for Computing Machinery, pp. 1215-1224.

Liu, J., Shang, M., & Chen, D. (2009). Personal recommendation based on weighted bipartite networks. *Sixth International Conference on Fuzzy Systems and Knowledge Discovery*.

Lorrain, F., & White, H. (1971). Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology*, Vol. 1, pp. 49-80.

McDonald, J.H. (2009a). Spearman rank correlation. *Handbook of Biological Statistics*, 2nd ed. Baltimore, MD: Sparky House Publishing; available online: http://udel.edu/~mcdonald/statspearman.html.

McDonald, J.H. (2009b). Multiple comparison. *Handbook of Biological Statistics*, 2nd ed. Baltimore, MD: Sparky House Publishing; available online: http://udel.edu/~mcdonald/statmultcomp.html.

Newman, M.E.J. (2001). Scientific collaboration networks. II. Shortest Paths, weighted networks, and centrality. *Physical Review E*, Vol. 64, 016132.

Newman, M.E.J. (2004). Analysis of weighted networks. *Physical Review E*, Vol. 70, 056131.

Opsahl, T. (2009). *Structure and equivalence of weighted networks*. London: UK: University of London, Queen Mary College, pp. 104-122; available online: http://toreopsahl.com/publications/thesis/ or http://toreopsahl.com/tnet/.

Opsahl, T. (2013). Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks*, Vol. 35, No. 2, pp. 159-167. Data available online: http://toreopsahl.com/datasets/#online_forum_network.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The pagerank citation ranking: Bringing order to the web. *Technical report, Stanford Digital Library Technologies*.

Padrón, B., Nogales, M., & Traveset, A. (2011). Alternative approaches of transforming bimodal into unimodal mutualistic networks: The usefulness of preserving weighted information. *Basic and Applied Ecology*, Vol. 12, pp. 713-721.

Pan, Y., Li, D., and Liang, J. (2010). Detecting community structure in complex networks via node similarity. *Physica A: Statistical Mechanics and its Applications*, Vol. 389, No. 14, pp. 2849-2857.

Ramsey, P.H. (1989). Critical values for Spearman's rank order correlation. *Journal of Educational Statistics*, Vol. 14, No. 3, pp. 243-253.

Valente, T.W. & Foreman, R.K. (1998). Integration and radiality: Measuring the extent of an individual's connectedness and reachability in a network. *Social Networks*, Vol. 20, pp. 89-109.

Valente, T.W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How correlated are network centrality measures? *Connections*, Vol. 28, No. 1, pp. 16-26.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York, NY: Cambridge University Press.

Zhou, T., Ren, J., Medo, M., & Zang, Y. (2007). Bipartite network projection and personal recommendation. *Physical Review E*, Vol. 76, 046115.