

A Family of Affiliation Indices for Two-Mode Networks*

Frank Tutzauer

Department of Communication, University at Buffalo, comfrank@buffalo.edu

Abstract

An affiliation network consists of actors and events. Actors are affiliated with each other by virtue of the events they mutually attend. This article introduces a family of affiliation measures that captures the extent of actors' affiliations in the network. At one extreme, one might have an actor who attended many events, but none of these events were attended by any of the other actors in the network. Although of high degree, in no reasonable interpretation would such an actor be considered highly affiliated with other actors in the network. At the other extreme, one might have an actor defined by a collection of events, all of which were attended by another actor(s), making the actor as enmeshed in the network as possible. Most actors will be between these extremes, with some events being shared by varying others, and some not. This article introduces a family of affiliation measures based on the entries of the co-occurrence matrix. After defining the measures, the cumulative distribution function of first-order affiliation is derived and expressed as a difference of binomials.

Keywords

Affiliation networks, network methodology, statistical methods, two-mode networks

*A previous version of this paper was presented at Sunbelt XXXI, the annual conference of the International Network for Social Network Analysis, St. Pete Beach, FL, February 8-13, 2011.

Introduction

Two-mode networks, in which the nodes of a network are partitioned into two groups, or modes, have received increasing attention in the social network literature (e.g., Latapy, Magnien and Del Vecchio 2008; Wang, Sharpe, Robins and Pattison 2009). The type of two-mode networks considered in this article are often called *affiliation networks*, and we think of one of the modes as a set of *actors* and the other mode as a set of *events*. Actors are affiliated with each other by virtue of the events they mutually attend. A classic example comes from Davis, Gardner and Gardner (1941) in which the actors are 18 southern women and the events are 14 social gatherings attended by varying numbers of the women. Even though one uses the terms *actors* and *events*, the objects of study need not be literal actors and events. For example, in the affiliation network of Ferrer i Cancho and Solé (2001), the “actors” are words and the “events” are the sentences in which they appear.

Affiliation networks have been used to study subjects as diverse as interlocking corporate directorates (Koenig, Gogel and Sonquist 1979), community organizations (Crowe 2007), director affiliation through both corporate and noncorporate organizations (Barnes and Burkett 2010), academic collaboration (Moody 2004), political communication (Chung and Park 2010; Park and Thelwall 2008), human language (Ferrer i Cancho and Solé 2001; Zhou and Heineken 2009), computer-mediated communication (Cho and Lee 2008), and the disciplinary structure of academic organizations (Barnett and Danowski 1992; Chung, Lee, Barnett and Kim 2009; Doerfel and Barnett 1999; Lee 2008).

Every affiliation network has a representation as a bipartite graph, where an edge is placed between an actor and an event only if the actor attended that event. Because certain edges are prohibited (namely edges between actors and edges between events), affiliation networks present certain problems (and opportunities) when one attempts to apply standard network measures (Borgatti and Everett 1997). This article develops a family of affiliation indices specifically for two-mode networks that are represented by bipartite graphs. I also derive the cumulative distribution function of one of the indices from the family and illustrate the measures on both hypothetical and empirical networks.

The structure of the paper is as follows: First, I introduce basic mathematical terminology. Then I define a family of affiliation indices and derive the sampling distribution of one of these indices. I then illustrate the ideas on an empirical network, and offer some concluding thoughts.

Mathematical Preliminaries

The networks in this article are loopless and undirected. Every such network has a mathematical representation as a graph $G = (V, L)$, where V is a finite, nonempty set of *vertices* or *nodes* and L is an irreflexive, symmetric relation on V . The elements of L are called *edges* or *links*. If $(v, w) \in L$, then the nodes v and w are *neighbors* and the edge (v, w) is *incident* with the nodes. The *degree* of a node v , symbolized $d(v)$, is the number of neighbors it has (with the parenthetical material dropped when the node is understood). In a two-mode network, the nodes can be partitioned into two groups meaningful to the investigator; for example, men and women or authors and their papers. Most often, it is assumed that every edge is incident with exactly one node from each group. In such a case, the representation is a bipartite graph. The graph $G = (V, L)$ is *bipartite* if V can be partitioned into two nonempty, disjoint subsets $A = \{v_1, v_2, \dots, v_n\}$ and $E = \{e_1, e_2, \dots, e_m\}$ such that every edge is incident with exactly one node from A and exactly one node from E . Two-mode networks with a bipartite representation are called *affiliation networks*, and we think of A as a set of *actors* and E as a set of

events. All edges are incident with exactly one actor and exactly one event. If an edge is incident with $v_i \in A$ and $e_j \in E$ we will say that actor v_i attends event e_j , and if two actors attend the same event we will say that they *share* the event.

Each bipartite graph gives rise to an $n \times m$ *affiliation matrix*, $\mathbf{A} = [a_{ij}]$, where $a_{ij} = 1$ if and only if the i th actor attends the j th event. The degree of v_i is given by the i th row sum of \mathbf{A} . If one post-multiplies the affiliation matrix by its transpose, one obtains the *co-occurrence matrix* $\mathbf{C} = \mathbf{A}\mathbf{A}'$. If $i \neq j$ then the ij th entry of \mathbf{C} is the number of events shared by actors v_i and v_j ; it is the degree of v_i if $i = j$.

How Should We Measure Affiliation?

The affiliation index of an actor should measure the extent to which the actor is fully integrated into the structure of the affiliation network—how enmeshed the actor is, by virtue of the actor's shared events, relative to other actors. The one-mode notion of centrality immediately comes to mind as a possible proxy, but upon reflection it doesn't take into account the richer types of connections possible in a two-mode network. (For applications of various centrality measures to two-mode networks see Borgatti and Everett 1997, and Faust 1997.) To see what I mean, consider Figures 1 and 2, which show two depictions of a 10-actor, 25-event affiliation network, and consider degree centrality, which is given by the degree of the actor. Inspection of Figure 1 shows that the most central actors are E, with a degree of 9, and actors G and H, each with a degree of 8. The extent of their integration into the network is quite different, however, though it is not obvious in the figure.

We can gain a little better insight by using a different visual depiction of the network. Figure 2 shows a hypergraph that corresponds to the bipartite graph of Figure 1. The mathematical exposition in the remainder of this article does not use hypergraphs, so there is no need to define them formally. It suffices at this point to say that in Figure 2 the actors are represented by the various geometric shapes and the events are represented by the numbers contained in the shapes. (The actual geometric shapes—triangles, rectangles, ovals, etc.—have no significance apart from the events they enclose. The different shapes are used only to help the eye distinguish among the various actors; had only a single style been used—rectangles, say—it would have been difficult visually to tell which actors attended which events.) Note that there is a one-to-one correspondence between the bipartite network of Figure 1 and the hypergraph of Figure 2. For example, in Figure 1 actor A is linked to events 1, 6, and 7. In the hypergraph, the triangle representing actor A encloses precisely these three events. Similarly, Figure 1 shows that actor J attended events 15 and 25, and in the hypergraph the upward sloping oval representing actor J encloses these two events. The degree of an actor and all other structural features are thus maintained in the two representations.

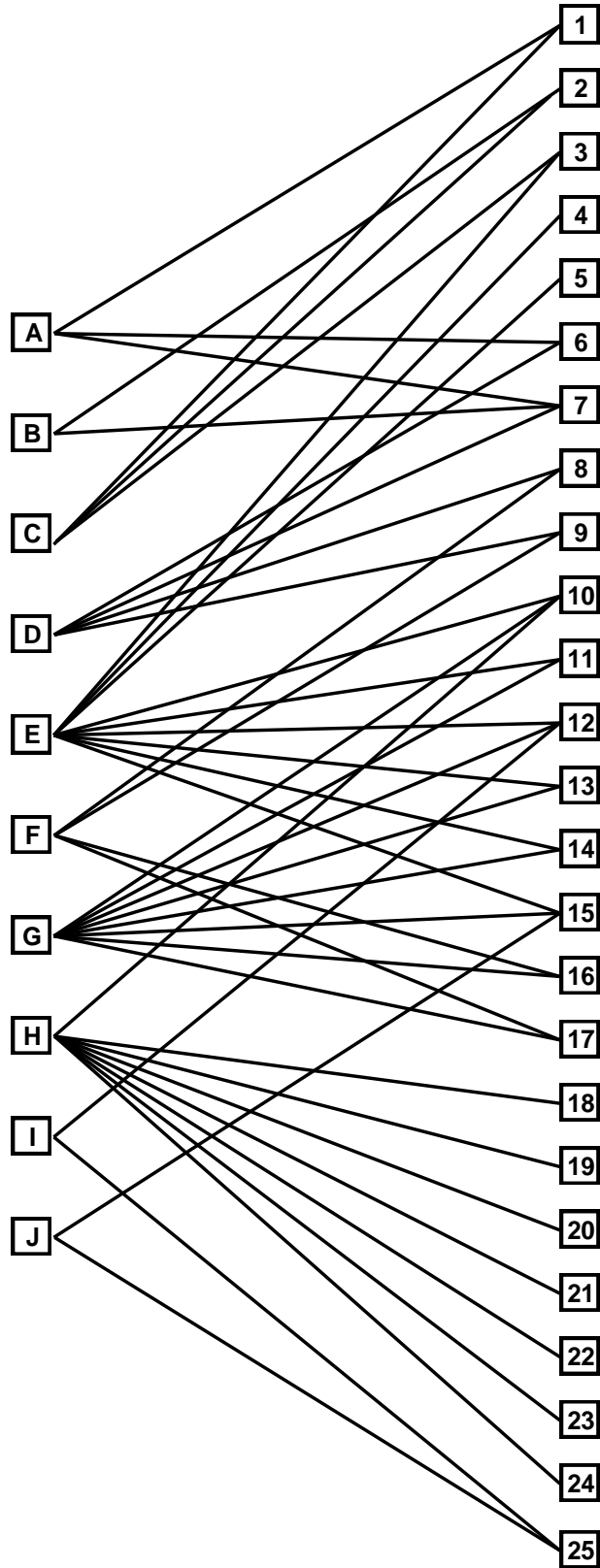


Figure 1. Bipartite graph representing a hypothetical affiliation network.

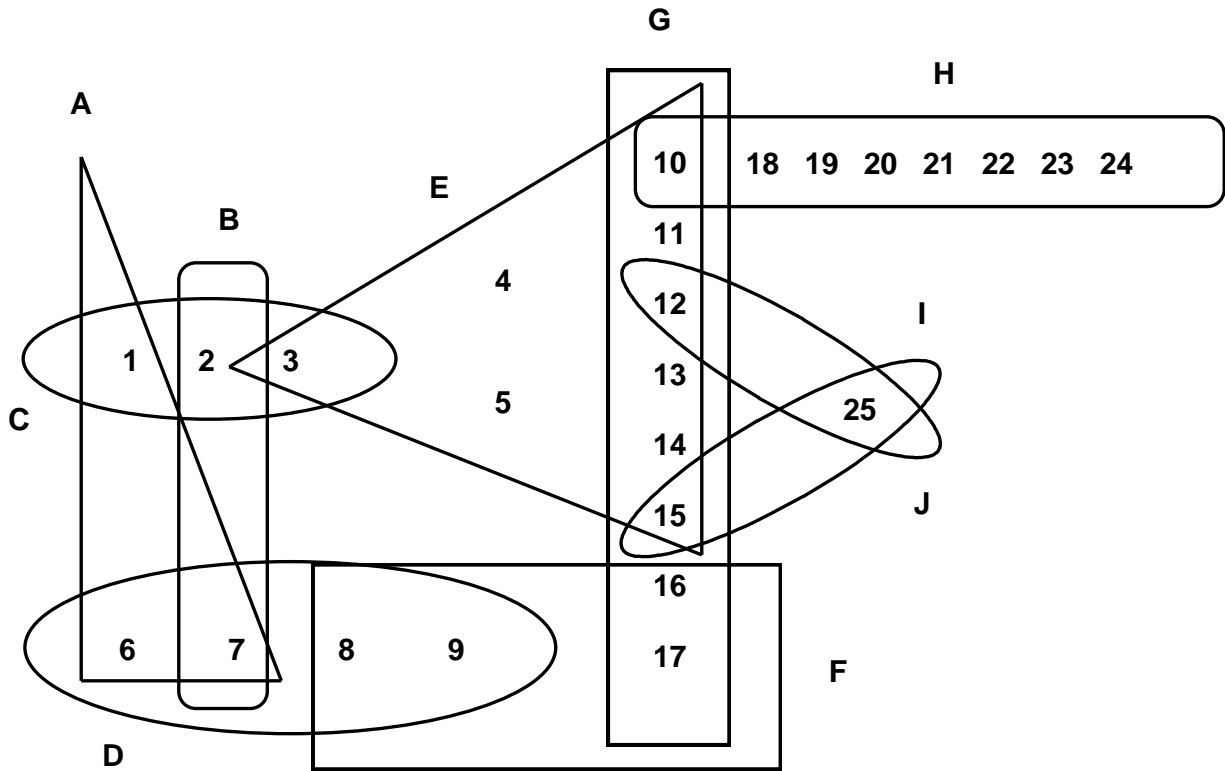


Figure 2. Hypergraph representing a hypothetical affiliation network.

Mathematically, we will work with the bipartite network and its affiliation matrix, but, visually, it is easier to see the shared affiliations in the hypergraph. The hypergraph also shows that the degree of an actor does not necessarily correspond to how affiliated the actor is with other actors. For example, as can be seen in Figure 2, actor H, though of high degree, is actually quite marginal in terms of shared events, attending only one event in common with other actors, namely event 10, which is shared by actors E and G. Actor G's affiliations, on the other hand, are quite a bit more extensive. All of the events attended by this actor were also attended by others, primarily actor E, but also actors F, H, I, and J. Actor E, with the highest degree, has nearly as many mutually attended events as G, but there are two events (4 and 5) not attended by other actors. In fact, in some networks centrality and affiliation might even be inversely related. From a purely probabilistic viewpoint, an actor of high degree, say, would be unlikely to share a large proportion of events with another actor, whereas an actor of low degree would be much more likely to have a large proportion of mutually attended events.

There are several ways in which one might assess an actor's affiliation using shared events. Their number, size, and multiplicity could be taken into account. As a first attempt, however, let us simply use the largest number of events that an actor shares with another (single) actor. We can then divide this number by the degree of the actor, giving us a proportion.

To formalize these notions, let $C = [c_{ij}]$ be the co-occurrence matrix and define the *mutual affiliation* of the distinct actors v_i and v_j to be the total number of events they share, which, of course, is given by c_{ij} . The *maximum mutual affiliation* of an actor v_i is defined to be $\mu(v_i) = \max\{c_{ij}\}$, where the maximum is

to be taken across all j not equal to i . In the co-occurrence matrix, $\mu(v_i)$ is the largest off-diagonal entry in the i th row. Using these notions, we define actor v_i 's *affiliation index*, symbolized $\alpha(v_i)$, by the equation:

$$\alpha(v_i) = \frac{\mu(v_i)}{d(v_i)}. \tag{1}$$

As before, the parenthetical material can be dropped when the actor is understood. Table 1 shows the degrees, maximum mutual affiliations, and the affiliation indices for each of the actors in Figures 1 and 2.

Table 1. *Degrees, Maximum Mutual Affiliations, and Affiliation Indices*

Actor	d	μ	α
A	3	2	.667
B	2	1	.500
C	3	1	.333
D	4	2	.500
E	9	6	.667
F	4	2	.500
G	8	6	.750
H	8	1	.125
I	2	1	.500
J	2	1	.500

Note: Actors are those in Figures 1 and 2. The symbols d , μ , and α refer to, respectively, the actors' degrees, maximum mutual affiliation scores, and affiliation indices.

As a measure of affiliation, α is not perfect. For example, in Figure 2, the maximum mutual affiliation for actor E consists of the 6 events shared with actor G. Not taken into account are the lone event shared with actor C and the fact that three events—10, 12, and 15—are shared with others in addition to actor G. In fact, one disadvantage of the measure is that it is somewhat “coarse.” To see what I mean, consider once again Figure 2 and focus on actors D and F. Both are of degree 4 and both have a maximum mutual affiliation of 2, giving an affiliation index of .500. One could certainly argue, however, that actor D is more embedded in the network than F. These actors share two events with each other, and each shares an additional two events with another actor (A for actor D and G for actor F), but actor D also shares an event with a third actor (actor B) whereas F does not. Because the present measure depends primarily on only an actor's *maximum* mutual affiliation without taking into account other mutual affiliations, actors D and F receive the same scores with the present measure whereas a finer-grained measure might distinguish between them.

The above analysis suggests that we might consider the present measure merely the first in a family of measures, what we might call the *first-order affiliation index* and symbolize α_1 . If we let μ_i be the i th largest mutual affiliation an actor has, then we would rewrite Equation 1 as $\alpha_1 = \mu_1/d$ (if the actor is understood). Moreover, a reasonable second-order affiliation index would be $\alpha_2 = (1/2)[(\mu_1 + \mu_2)/d]$. In general, for $k = 1, \dots, n - 1$ we define the k th-order affiliation index for an actor to be the sum of the actor's k largest mutual affiliations all divided by kd (dividing by k normalizes the measure to a 0-1 scale). For actors D and F of Figure 2, the first- and second-order affiliation indices do not distinguish between them, but the third-order affiliation index does, giving actor D a score of .417 and actor F a score of .333, according with our insight that actor D is more embedded in the network than actor F.

Table 2 shows affiliation indices of all orders for the actors in Figures 1 and 2. In addition to distinguishing between actors D and F, three actors—B, I, and J—are seen to have identical patterns of affiliation. Each of these actors attends two events, one shared by two others and one shared by just one other, and so they have identical affiliation indices for all orders.

Table 2. *Affiliation indices of all orders for hypothetical network*

Order									
Actor	1	2	3	4	5	6	7	8	9
A	0.667	0.500	0.444	0.333	0.267	0.222	0.190	0.167	0.148
B	0.500	0.500	0.500	0.375	0.300	0.250	0.214	0.188	0.167
C	0.333	0.333	0.333	0.250	0.200	0.167	0.143	0.125	0.111
D	0.500	0.500	0.417	0.313	0.250	0.208	0.179	0.156	0.139
E	0.667	0.389	0.296	0.250	0.222	0.185	0.159	0.139	0.123
F	0.500	0.500	0.333	0.250	0.200	0.167	0.143	0.125	0.111
G	0.750	0.500	0.375	0.313	0.275	0.229	0.196	0.172	0.153
H	0.125	0.125	0.083	0.063	0.050	0.042	0.036	0.031	0.028
I	0.500	0.500	0.500	0.375	0.300	0.250	0.214	0.188	0.167
J	0.500	0.500	0.500	0.375	0.300	0.250	0.214	0.188	0.167

Note: Actors are those in Figures 1 and 2.

In practice, it is probably not necessary to compute affiliation indices of all orders. In fact, first-order affiliation may be sufficient in many applications. Moreover, the first-order affiliation index has three advantages: First, it is intuitive; it is equal to the largest proportion of events that an actor attends with another actor. Second, it is easy to compute; the maximum mutual affiliations can be taken from the co-occurrence matrix and the degrees can be taken from either the co-occurrence matrix or the affiliation matrix. Finally, the measure leads to a mathematically tractable derivation of its sampling distribution, a topic that is taken up next.

Sampling Distribution of First-Order Affiliation

Statistical Model

Throughout this section we deal only with first-order affiliation, so we will drop the subscripts from α and μ . In other words, as in Equation 1, α will stand for the first-order affiliation index and μ will stand for the actor's maximum mutual affiliation. With this notational convenience established, we start the derivation by assuming that the n actors and m events are fixed, and that the ij th entry of the $n \times m$ affiliation matrix \mathbf{A} is 1 with fixed and independent probability p and 0 with probability $1 - p$. Such a statistical model was first studied in the one-mode case by Gilbert (1959) and by the more widely cited Erdős and Rényi (1960). It serves as a useful baseline from which to assess whether or not observed features in an empirical network differ from what would be expected by chance. As Alderson (2008) has said, this model serves "as a natural 'null hypothesis' for evaluating properties of network structure" (p. 1051). For example, Aviv, Erlich and Ravid (2008) compared online distance learning networks to an Erdős-Rényi baseline. Kang (2007) used this statistical model to explore the relationship between degree equicentrality and degree centralization, Mizruchi and Neuman (2008) used it to examine bias in network autocorrelation models, and Field, Frank, Schiller, Riegle-Crumb and Muller (2006) used it to generate the sampling distribution of their clustering coefficient in two-mode data. Ko, Lee and Park (2008) used this statistical model to explore degree and closeness centrality, and Faust (2008) used it to generate the distribution of the triad census. Finally, Chung and Lu (2001) have noted its application to the Internet.

Under the assumption that the elements of the affiliation matrix are generated independently, many of the variables of interest will have a binomial distribution. So that it can be used in the sequel, we define it now. Let X be a discrete, random variable. Let v be a nonnegative integer and let p be a real number such that $0 \leq p \leq 1$. Then X has a binomial distribution with parameters v and p if, for any nonnegative integer $k \leq v$, the probability that X equals k is given by the probability mass function

$$b(k) = P(X = k) = \binom{v}{k} p^k (1 - p)^{v-k}. \quad (2)$$

To avoid cluttering the notation, we will not make the dependence of b on its parameters explicit in the symbols, but rather will simply state what the parameters are if it is not already apparent.

When p is the probability that the ij th entry of \mathbf{A} is 1, and when $v = m$, then, for actor v_i , $b(k)$ is the probability that this actor attends exactly k events—that is, $b(k)$ is the probability that $d(v_i) = k$. The associated cumulative distribution function (cdf) is

$$B(k) = P(X \leq k) = \sum_{i=0}^k \binom{v}{i} p^i (1 - p)^{v-i}. \quad (3)$$

If X is distributed other than as a binomial, we will denote its probability mass function by f_X and its cumulative distribution function by F_X .

Derivation of the Distribution

Consider the n th and final actor—actor v_n with degree $d(v_n)$. We do so to simplify notation. If instead we choose, say, the i th actor, then many of our summations would have to be split into two: one summation for actors 1 through $i - 1$ and one for actors $i + 1$ to n . It is notationally simpler to focus on the n th actor and have a single summation running from 1 to $n - 1$. Note, however, that the argument below is completely general; once we have derived the desired sampling distribution for the maximum mutual affiliation of v_n , the same formulae work for any actor (of course, using that actor's degree rather than v_n 's).

Without loss of generality, then, consider actor v_n . If this actor is of very low degree then it is highly likely that $\mu(v_n)$ is very close to $d(v_n)$. Alternatively, if $d(v_n)$ is very large, then it is extremely unlikely that $\mu(v_n)$ is very near zero, on the one hand, or very near $d(v_n)$, on the other. Most likely it is somewhere in the middle. Thus, the degree of an actor influences in large part the actor's maximum mutual affiliation. Consequently, we will proceed by conditioning on degree and deriving the distribution of μ , from which we will be able to deduce likelihood of a given actor's first-order affiliation.

The first question to ask is how many events v_n shares with another actor v_i , $i \neq n$. There are exactly $d(v_n)$ columns in the n th row of the affiliation matrix that are equal to 1 (with the rest being equal to 0). If we look in these same $d(v_n)$ columns in the i th row and call 1 a "success" and 0 a "failure" then we have a sequence of $d(v_n)$ Bernoulli trials with probability of success equal to p . Now let Q be the number of the $d(v_n)$ events shared by v_n and v_i . Q is distributed as a binomial with parameters $d(v_n)$ and p :

$$P(Q = k) = b(k). \quad (4)$$

What we are really after though is whether or not $\mu(v_n) = k$, that is whether or not the largest number of events that v_n shares with another actor is k . Focusing again on v_i , three outcomes are of significance: whether v_n and v_i share fewer than k events, exactly k events, or more than k events. The probabilities of these three outcomes are:

$$p_1 = P(Q < k) = \sum_{i=0}^{k-1} \binom{d(v_n)}{i} p^i (1-p)^{d(v_n)-i} = B(k-1), \quad (5)$$

$$p_2 = P(Q = k) = \binom{d(v_n)}{k} p^k (1-p)^{d(v_n)-k} = b(k), \quad (6)$$

and,

$$p_3 = P(Q > k) = \sum_{i=k+1}^{d(v_n)} \binom{d(v_n)}{i} p^i (1-p)^{d(v_n)-i} = 1 - B(k), \quad (7)$$

where B is the cumulative distribution of the binomial with parameters $d(v_n)$ and p , and where b is its probability mass function.

Now, consider the other $n - 1$ actors and their mutual affiliations with v_n . Let Q_1 be the number of actors having a mutual affiliation with v_n of less than k , let Q_2 be the number with a mutual affiliation of exactly k , and let Q_3 be the number with a mutual affiliation of more than k . The likelihood that $Q_i = k_i$, for $i = 1, 2, 3$, is the multinomial probability

$$P(Q_1 = k_1, Q_2 = k_2, Q_3 = k_3) = \frac{(n-1)!}{k_1!k_2!k_3!} p_1^{k_1} p_2^{k_2} p_3^{k_3}. \quad (8)$$

Thus, to obtain the probability that $\mu(v_n) = k$, we merely need to note each of the ways that this result can occur, weight by the probability given in Equation 8, and add. The possible values of Q_1, Q_2 , and Q_3 that would result in $\mu(v_n) = k$ are shown in Table 3. Examining the first row of the table we see that $\mu(v_n)$ would equal k if exactly 1 actor shared k events with v_n and the other $n - 2$ actors shared fewer than k events with v_n . Similarly, from the second row, if there are exactly 2 actors that share k events with v_n and $n - 3$ that share fewer than k events then the maximum mutual affiliation would again equal k . Table 3 was constructed by noting, first, that Q_3 must always be equal to zero (because if an actor shares more than k events with v_n then k is not maximum) and by noting, second, that Q_2 must always be greater than or equal to 1 (and thus the remaining actors must all share fewer than k vertices with v_n).

Table 3. Ways in Which $\mu(v_n) = k$ and Their Associated Probabilities

Q_1	Q_2	Q_3	Probability
$n - 2$	1	0	$P(n - 2, 1, 0)$
$n - 3$	2	0	$P(n - 3, 2, 0)$
$n - 4$	3	0	$P(n - 4, 3, 0)$
.	.	.	.
.	.	.	.
.	.	.	.
1	$n - 2$	0	$P(1, n - 2, 0)$
0	$n - 1$	0	$P(0, n - 1, 0)$

Note: Q_1, Q_2, Q_3 are, respectively, the number of actors with less than, equal to, or greater than k shared events with actor v_n . The notation $P(x, y, z)$ in the fourth column is shorthand for $P(Q_1 = x, Q_2 = y, Q_3 = z)$ as in Equation 8.

Adding up all the ways that the maximum mutual affiliation could equal k and weighting by the associated probability in Table 3, we arrive at the probability mass function:

$$f_\mu(k) = P(\mu = k) = \sum_{i=1}^{n-1} \frac{(n-1)!}{(n-1-i)!i!0!} p_1^{n-1-i} p_2^i p_3^0. \quad (9)$$

Equation 9 gives the probability that an actor with a specified degree has a maximum mutual affiliation of exactly k . We can simplify Equation 9 somewhat. First note that p_3^0 as well as $0!$ in the denominator of the fraction are both equal to 1, and thus drop out. Then note that the remaining fraction is the

number of combinations of $n - 1$ objects taken i at a time. Thus we obtain

$$f_{\mu}(k) = \sum_{i=1}^{n-1} \binom{n-1}{i} p_1^{n-1-i} p_2^i. \quad (10)$$

We can make further progress by adding and subtracting a term for $i = 0$. Doing so lets us run the summation from 0 to $n - 1$, putting it in a form that allows us to use the Binomial Theorem:

$$\begin{aligned} f_{\mu}(k) &= \left[\sum_{i=1}^{n-1} \binom{n-1}{i} p_1^{n-1-i} p_2^i + \binom{n-1}{0} p_1^{n-1-0} p_2^0 \right] - \binom{n-1}{0} p_1^{n-1-0} p_2^0 \\ &= \left[\sum_{i=0}^{n-1} \binom{n-1}{i} p_1^{n-1-i} p_2^i \right] - p_1^{n-1} \\ &= (p_1 + p_2)^{n-1} - p_1^{n-1} \\ &= [B(k-1) + b(k)]^{n-1} - B(k-1)^{n-1}, \end{aligned} \quad (11)$$

where the third line of Equation 11 follows from the second by applying the Binomial Theorem to the expression in brackets, and where the fourth line follows from the third by using the definitions of the p 's in Equations 5 and 6. Recognizing that $B(k) = B(k-1) + b(k)$, we see that the probability mass function can be conveniently expressed as a difference of (exponentiated) binomial cdfs

$$f_{\mu}(k) = B(k)^{n-1} - B(k-1)^{n-1}. \quad (12)$$

The cumulative distribution function is

$$F_{\mu}(k) = \sum_{i=0}^k [B(i)^{n-1} - B(i-1)^{n-1}]. \quad (13)$$

The parameters of the binomial cdfs in Equations 12 and 13 are the degree of the actor and the baseline probability of event attendance, p , which in most applications would be the density of the affiliation matrix (i.e., $p = (nm)^{-1} \Sigma a_{ij}$).

We will say that a random variable has an $\mathcal{A}(n, d, p)$ distribution if its cdf is given by Equation 13. With respect to affiliation networks, if we take an actor's first-order affiliation index and multiply through by the degree of the actor, then the resulting test statistic (μ) has an $\mathcal{A}(n, d, p)$ distribution, where n is the number of actors in the network, d is the degree of the actor, and p is the density of the affiliation matrix.

An actor with a very low affiliation score will have a percentile in the left-hand tail of the distribution. In particular, an actor with maximum mutual affiliation equal to k will have a first-order affiliation index of $\alpha = k/d$ or lower with probability

$$P(\mu \leq k) = F_{\mu}(k). \quad (14)$$

If this probability is less than .05 (or whatever alpha level the researcher selects), then the actor's affiliation score is significantly less than would be expected by chance. Alternatively, for large affiliation indices, the appropriate right-hand tail probability is

$$P(\mu \geq k) = 1 - F_{\mu}(k - 1). \quad (15)$$

If this probability is less than the selected alpha level, then the actor has a first-order affiliation score significantly higher than expected by chance.

Because they are based on binomial cdfs, the percentiles of the $\mathcal{A}(n, d, p)$ distribution are straightforward to calculate. Tables of binomial cdfs are given in many standard statistics texts (e.g., Guenther 1973, and Wonnacott and Wonnacott 1990), and there are many fast and accurate cumulative binomial calculators on the Internet. Applying Equations 15 and 16 to the hypothetical data of Figures 1 and 2, we find three significant results. First, actor H with a degree of 8 and a maximum mutual affiliation of 1 has an affiliation index significantly less than expected by chance. The likelihood of an actor with this degree having such a low affiliation score is .0057, which is easily significant at the .01 level, one-tailed. Second, actor G also has a degree of 8 but a maximum mutual affiliation of 6. The probability of an actor with degree 8 attending 6 or more events is .0061, again significant at the .01 level, one-tailed. Finally, actor E has a degree of 9 and a maximum mutual affiliation of 6, which leads to a righthand tail probability of .0155, significant at the .05 level, but not the .01 level. Doing similar computations for the rest of the actors, we see that none of them have significant first-order affiliation scores. The conclusion, then, is that actor H is significantly less affiliated than expected by chance, and actors E and G are significantly more affiliated than expected by chance; the remaining actors have first-order affiliation scores that are statistically the same as if events were attended at random.

An Empirical Example

In this section, we apply the first-order affiliation index and its sampling distribution to an empirical, two-mode network. Consider Table 4, which shows the affiliation matrix of 20 corporate directors (the actors) and their memberships in social clubs and various corporate, museum, and university boards of directors or trustees (the events), as reported by Barnes and Burkett (2010). There are 99 edges in this network, giving a density of .21. Table 5 shows the co-occurrence matrix, and Table 6 shows the first-order affiliation scores of the directors along with various two-mode centrality scores, normalized to a 0-1 scale. UCINET 6 (Borgatti, Everett and Freeman 2002), which was used to compute the centralities, adjusts the denominators of the centrality measures for two-mode networks compared to one-mode networks. For example, to normalize degree centrality, one divides the degree by the maximum possible degree, which in a one-mode network is one fewer than the number of vertices in the network. In an affiliation network, the appropriate denominator is the number of events because actors cannot be neighbors of other actors. UCINET makes similar adjustments for the other centrality measures as well. (For details, see Borgatti and Everett 1997.)

Table 4. Affiliation Matrix of 20 Corporate Directors

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Barr	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	1	0	0
Block	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0
Chewning	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Clark	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
Cushman	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Eberhard	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Freeman	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
Gale	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	1	1	0	0	1	0	1	0
Goodrich	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	1	0	0
Ingersoll	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
Jarvis	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	1	0	1	0	0	1	0	0
Kennedy	0	0	0	0	1	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	1	1	0	0
Livingston	1	0	0	0	0	0	0	0	1	1	0	0	1	1	0	1	0	0	1	0	1	1	0	0
McCormick	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	0	0	0	0	1	1	0	0
McDowell	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
Oates	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	1	1	0	1	1	1	0	1
Prince	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Rockefeller	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1
Swearingen	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Ward	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0	0

Note: Events are numbered as follows. Corporate Boards: 1 = Amour, 2 = Borg-Warner, 3 = Caterpillar Tractor, 4 = Chase Manhattan Bank, 5 = Commonwealth Edison, 6 = Container Corp. of America, 7 = Continental International Bank and Trust, 8 = Equitable Life Assurance, 9 = First National Bank of Chicago, 10 = Inland Steel, 11 = International Harvester, 12 = John Hancock Mutual, 13 = Sears and Roebuck, 14 = Standard Oil, 15 = Swift. Museums: 16 = Art Institute of Chicago, 17 = Museum of Science and Industry. University Boards: 18 = Northwestern Univ., 19 = Univ. of Chicago. Social Clubs: 20 = Century, 21 = Chicago, 22 = Commercial, 23 = Indian Hill, 24 = Links.

Table 5. Co-Occurrence Matrix for 20 Corporate Directors

		1	2	3	4	5	6	7	8	9	0	1	1	1	1	1	1	1	1	1	2
		1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0
1	Barr	5	2	0	1	0	1	2	3	3	2	3	4	2	3	2	3	0	0	0	3
2	Block	2	5	0	0	1	1	2	3	1	2	2	2	3	2	2	3	1	0	1	2
3	Chewning	0	0	2	2	0	0	0	0	0	0	0	0	1	0	0	0	2	0	0	0
4	Clark	1	0	2	3	0	1	1	1	1	1	0	1	2	1	0	1	2	0	0	1
5	Cushman	0	1	0	0	2	1	1	0	0	1	0	0	2	1	1	1	1	0	1	0
6	Eberhard	1	1	0	1	1	3	3	1	1	2	0	1	2	2	1	2	1	0	1	1
7	Freeman	2	2	0	1	1	3	6	1	2	5	1	2	3	3	2	3	1	0	1	2
8	Gale	3	3	0	1	0	1	1	7	1	2	2	3	2	1	0	3	0	0	0	2
9	Goodrich	3	1	0	1	0	1	2	1	4	2	3	4	2	4	2	2	0	0	0	3
10	Ingersoll	2	2	0	1	1	2	5	2	2	6	1	2	3	3	2	3	1	0	1	2
11	Jarvis	2	2	0	0	0	0	1	2	3	1	6	5	2	3	2	2	0	1	0	3
12	Kennedy	4	2	0	1	0	1	2	3	4	2	5	7	3	4	2	2	0	1	0	5
13	Livingston	2	3	1	2	2	2	3	2	2	3	2	3	9	4	2	3	2	1	2	3
14	McCormick	3	2	0	1	1	2	3	1	4	3	3	4	4	6	3	3	1	0	1	3
15	McDowell	2	2	0	0	1	1	2	0	2	2	2	2	2	3	3	2	1	0	1	1
16	Oates	3	3	0	1	1	2	3	3	2	3	2	2	3	3	2	9	1	4	2	2
17	Prince	0	1	2	2	1	1	1	0	0	1	0	0	2	1	1	1	3	0	1	0
18	Rockefeller	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	4	0	5	1	1
19	Swearingen	0	1	0	0	1	1	1	0	0	1	0	0	2	1	1	2	1	1	3	0
20	Ward	3	2	0	1	0	1	2	2	3	2	3	5	3	3	1	2	0	1	0	5

Table 6. *First-Order Affiliation and Two-Mode Centrality for 20 Corporate Directors*

Director	Affiliation	Degree	Closeness	Betweenness	Eigenvector
Barr	0.800	0.208	0.564	0.024	0.261
Block	0.600	0.208	0.585	0.034	0.224
Chewning	1.000	0.083	0.373	0.002	0.019
Clark	0.667	0.125	0.534	0.041	0.099
Cushman	1.000	0.083	0.492	0.006	0.070
Eberhard	1.000	0.125	0.564	0.019	0.150
Freeman	0.833*	0.250	0.608	0.078	0.263
Gale	0.429	0.292	0.574	0.061	0.213
Goodrich	1.000*	0.167	0.554	0.011	0.243
Ingersoll	0.833*	0.250	0.608	0.077	0.264
Jarvis	0.833*	0.250	0.554	0.034	0.242
Kennedy	0.714	0.292	0.596	0.051	0.330
Livingston	0.444	0.375	0.674	0.197	0.329
McCormick	0.667	0.250	0.608	0.050	0.319
McDowell	1.000	0.125	0.554	0.009	0.186
Oates	0.444	0.375	0.660	0.199	0.314
Prince	0.667	0.125	0.517	0.033	0.079
Rockefeller	0.800	0.208	0.456	0.023	0.075
Swearingen	0.667	0.125	0.508	0.017	0.083
Ward	1.000**	0.208	0.574	0.023	0.268

* $p < .05$, one-tailed; ** $p < .01$, one-tailed.

Examining the diagonal of Table 5, we see that the directors attended anywhere from a low of two events (Chewning and Cushman) to a high of nine events (Livingston and Oates). But attending many events does not mean that these events are shared with another actor. Examining Table 6, however, one sees that there are no actors with extremely low affiliation scores. The lowest proportion of an actor's events shared with another actor is just under 43 percent (Gale), while many of the directors share all of their events with at least one other actor.

Using Equations 14 and 15 to compute significance levels, one finds that none of the actors are significantly less affiliated than expected by chance. Five actors, however, are significantly more affiliated than expected by chance, four at the .05 level and one at the .01 level (all one-tailed). Two of them (Goodrich and Ward, $p = .034$ and $p = .007$, respectively) share all of their events with another actor, and the remaining three (Freeman, Ingersoll, and Jarvis, all with $p = .035$), each of degree six, share all save one of their events with another actor. Note that having a first-order affiliation of 1.00 does not assure significance. The degree of the actor also needs to be high enough that we can be confident that the shared events are not due to chance. For example, an actor of degree 2 in a 20-actor network with a density of .21, has a .562 probability of sharing both of those events with another actor, not something we would ordinarily call significant; it is just too easy for such an actor to reach a first-

order affiliation of 1.00. For an actor of degree 3, the probability of sharing all of those events with another actor drops to .154, and by the time we reach degree 4 the probability is low enough (.034) that we conclude that the observed affiliation score is not due to chance.

Note importantly that, as in the hypothetical network of the previous section, having a high degree does not necessarily mean that the actor will be highly affiliated in the network. In fact, none of the actors with the highest degrees (Livingston and Oates, each with degree 9, and Gale and Kennedy, each with degree 7) are significantly more affiliated than expected, although Kennedy is close ($p = .097$). Overall, the first-order affiliation scores fail to match the degree centralities as well as the other three centrality measures. In fact, as Table 7 shows, first-order affiliation is actually *negatively* correlated with the centrality measures for this network, a possibility remarked upon previously, and one that demonstrates that centrality and affiliation are quite different ideas. Whereas central actors may attend many events, or be between or close to other actors and events, affiliation specifically takes into account the nature of the *shared* events that actors have. Measuring affiliation is thus a purely two-mode idea. Centrality is, at its essence, a one-mode idea that we have tried to adapt (perhaps imperfectly) to the two-mode case; the notion of shared events does not even make sense in the one-mode case, and a measure of affiliation is thus most appropriate for two-mode data.

Table 7. *Correlation Matrix*

	Affiliation	Degree	Closeness	Betweenness	Eigenvector
Affiliation	---	-.669	-.503	-.695	-.334
Degree		---	.781	.832	.822
Close			---	.721	.884
Between				---	.578
Eigen					---

Discussion

This article has introduced a family of affiliation indices for two-mode networks having a bipartite representation. One of these indices—first-order affiliation, which is based on the largest proportion of events that an actor has in common with another actor in the network—was examined for its sampling characteristics. Using the Erdős-Rényi statistical model, the distribution of this measure was derived and was found to be conveniently expressible as a difference of binomial cdfs. One can use this distribution to test whether an actor's first-order affiliation is more or less likely than expected by chance.

In terms of future directions, because the first-order affiliation index can be somewhat coarse, some guidance is needed in selecting a higher-order affiliation index if one wants a finer-grained measure. It is a simple matter to compute affiliation indices of all orders, so this decision could be made on an ad hoc basis, though in a very large network, such a procedure would require the researcher to examine perhaps thousands of individual scores. An alternative would be to compute affiliation indices on a number of empirical networks to see if some pattern or rule of thumb can be developed to guide the decision of what order affiliation index to use.

A second direction of future research would be to derive the sampling distribution for higher order affiliation indices. Although it is not clear that a convenient, closed form expression can be found, it nonetheless might be possible to use previous results in order to obtain new results. For example, to derive the distribution of second-order affiliation, which depends on the sum $\mu_1 + \mu_2$, one could make an argument analogous to the one above to compute the conditional probability of μ_2 given μ_1 . This quantity, when combined with Equation 12, gives us the joint distribution of μ_1 and μ_2 . Whether such a procedure can be easily extended to other orders is an open question.

Until a convenient mathematical expression can be found, we can fall back on Monte Carlo analysis, in which one randomly generates a large number of affiliation matrices with the given row sums. Affiliation indices of various orders can be computed each time, giving a simulated distribution from which we can obtain probability values. Alternatively, a permutation approach could be used where actors are randomly assigned to events in ways that match event size, giving a comparison distribution.

Affiliation networks have wide applications in the social sciences, and, because of their special structure relative to traditional, one-mode networks, new tools and measures can fruitfully be developed—tools and measures that take advantage of two-mode information. Of particular interest is the way that actors are affiliated with other actors by virtue of their mutual attendance at events. This article has introduced a family of affiliation indices that captures, to varying degrees (depending on the order of the index), the nature of the mutual affiliations an actor has in the network. As more researchers develop tools, measures, and techniques such as those in this article, we approach the day when two-mode networks will be on an equal methodological footing with one-mode networks.

References

- Alderson, D. L. (2008). "Catching the "Network Science" Bug: Insight and Opportunity for the Operations Researcher." *Operations Research* 56, 5: 1047-1065.
- Aviv, R., Z. Erlich and G. Ravid (2008). "Analysis of Transitivity and Reciprocity in Online Distance Learning Networks." *Connections* 28, 1: 27-39.
- Barnes, R. and T. Burkett (2010). "Structural Redundancy and Multiplicity in Corporate Networks." *Connections* 30, 2: 4-20.
- Barnett, G. A. and J. A. Danowski (1992). "The Structure of Communication: A Network Analysis of the International Communication Association." *Human Communication Research* 19: 264-285.
- Borgatti, S. P. and M. G. Everett (1997). "Network Analysis of 2-Mode Data." *Social Networks* 19: 243-269.
- Borgatti, S. P., M. G. Everett and L. C. Freeman (2002). *UCINET for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.
- Cho, H. and J-S. Lee (2008). "Collaborative Information Seeking in Intercultural Computer-Mediated Communication Groups: Testing the Influence of Social Context Using Social Network Analysis." *Communication Research* 35: 548-573.
- Chung, C. J., S. Lee, G. A. Barnett and J. H. Kim (2009). "A Comparative Network Analysis of the Korean Society of Journalism and Communication Studies (KSJCS) and the International Communication Association (ICA) in the Era of Hybridization." *Asian Journal of Communication* 19: 170-191.
- Chung, C. J. and H. W. Park (2010). "Textual Analysis of a Political Message: The Inaugural Addresses of Two Korean Presidents." *Social Science Information* 49: 215-239.
- Chung, F. and L. Lu (2001). "The Diameter of Sparse Random Graphs." *Advances in Applied Mathematics* 26: 257-279.
- Crowe, J. A. (2007). "In Search of a Happy Medium: How the Structure of Interorganizational Networks Influence Community Economic Development Strategies." *Social Networks* 29: 469-488.
- Davis, A., B. B. Gardner and M. R. Gardner (1941). *Deep South: A Sociological Anthropological Study of Caste and Class*. Chicago, IL: University of Chicago.
- Doerfel, M. L. and G. A. Barnett (1999). "A Semantic Network Analysis of the International Communication Association." *Human Communication Research* 25: 589-603.
- Erdős, P. and A. Rényi (1960). "On the Evolution of Random Graphs." *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5: 17-61.
- Faust, K. (1997). "Centrality in Affiliation Networks." *Social Networks* 19: 157-191.

- Faust, K. (2008). "Triadic Configurations in Limited Choice Sociometric Networks: Empirical and Theoretical Results." *Social Networks* 30: 273-282.
- Ferrer i Cancho, R. and R. V. Solé (2001). "The Small world of Human Language." *Proceedings of the Royal Society B* 268: 2261-2265.
- Field, S., K. A. Frank, K. Schiller, C. Riegler-Crumb and C. Muller (2006). "Identifying Positions from Affiliation Networks: Preserving the Duality of People and Events." *Social Networks* 28: 97-123.
- Gilbert, E. N. (1959). "Random Graphs." *Annals of Mathematical Statistics* 30: 1141-1144.
- Guenther, W. C. (1973). *Concepts of Statistical Inference* (2nd ed.). New York: McGraw-Hill.
- Kang, S. M. (2007). "Equicentrality and Network Centralization: A Micro-Macro Linkage." *Social Networks* 29: 585-601.
- Ko, K., K. J. Lee and C. Park (2008). "Rethinking Preferential Attachment Scheme: Degree Centrality versus Closeness Centrality." *Connections* 28, 1: 4-15.
- Koenig, T., R. Gogel and J. Sonquist (1979). "Models of the Significance of Interlocking Corporate Directorates." *American Journal of Economics and Sociology* 38: 173-186.
- Latapy, M., C. Magnien and N. Del Vecchio (2008). "Basic Notions for the Analysis of Large Two-Mode Networks." *Social Networks* 30: 31-48.
- Lee, S. (2008). "Understanding the Nature of Citation: Examination of Structure in Citation Analysis and its Application to Communication Discipline in the Information Age." Doctoral dissertation, University at Buffalo, State University of New York, Amherst, NY.
- Mizruchi, M. S. and E. J. Neuman (2008). "The Effect of Density on the Level of Bias in the Network Autocorrelation Model." *Social Networks* 30: 190-200.
- Moody, J. (2004). "The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999." *American Sociological Review* 69: 213-238.
- Park, H. W. and M. Thelwall (2008). "Developing Network Indicators for Ideological Landscapes from the Political Blogosphere in South Korea." *Journal of Computer-Mediated Communication* 13: 856-879.
- Wang, P., K. Sharpe, G. L. Robins and P. E. Pattison (2009). "Exponential Random Graph (p*) Models for Affiliation Networks." *Social Networks* 31: 12-25.
- Wonnacott, T. H. and R. J. Wonnacott (1990). *Introductory Statistics* (5th ed.). New York: John Wiley.
- Zhou, D., and E. Heineken (2009). "The Use of Metaphors in Academic Communication: Traps or Treasures." *Ibérica* 18: 23-42.