Tracking Group Co-membership on Networks

James P. Ferry
Metron, Inc.
Reston, VA. U.S.A.
ferry@metsci.com

J. Oren Bumgarner
Metron, Inc.
Reston, VA. U.S.A.
bumgarner@metsci.com

Abstract - Tracking groups in network data is an emerging problem in network science. The network science community has not leveraged the tracking techniques used in data fusion, however. The purpose of this work is to introduce a novel domain to the tracking community, and novel techniques to network science. Group tracking is formulated here as a traditional, continuous-time Bayesian filter, which operates on time-evolving network data and outputs joint group membership probabilities over all nodes. Simple measurement and update models are proposed, which enable the derivation of an exact filter. This filter requires an exponentially large state space, however, so it is marginalized to a smaller space. The resulting system tracks second-order statistics (i.e., probabilities of pairs of nodes being in the same group) using equations involving third- and fourth-order statistics, which require closure assumptions. Several closures are investigated, and their merits and drawbacks are discussed.

Keywords: Tracking, networks, group finding, Bayesian filtering.

1 Introduction

Network data often exhibits groups of tightly connected nodes, representing, e.g., communities of people, clusters of related concepts, or sets of interacting proteins. A large number of algorithms for finding groups have been developed in the past eight years, including three particularly good ones in the past two [1]. Relatively little attention has been given to the problem of *tracking* groups on *dynamic* network data. What literature exists is covered in 3½ pages of a recent 100-page review of group finding [2]. Two important papers are [3], which studied the persistence of robust communities in the NEC CiteSeer database, and [4], which analyzed the evolution of overlapping groups in cell phone and co-authorship data.

These group tracking papers bear little relationship to the literature of conventional, kinematic tracking [5]. The purpose of this work is to make such a connection, recasting the group tracking problem in terms of a state space with motion and measurement models. This effort was begun in [6]. There it is assumed that the group membership of nodes and the states of links are each governed by independent, continuous-time Markov processes, where the transition rate matrices governing a link depend on the groups of its endpoints. For example, the transition rate from "link absent" to "link present" would typically be higher for node pairs in the same group than for pairs in different groups: this promotes higher link densities within groups than between them. The main result of [6] was an exact law governing the joint probability distribution of all nodes' group memberships given the time history of all the link states. This inference law was demonstrated on a 12-node example case—unfortunately the extremely large state space for the full joint distribution limits the applicability of the exact method to systems of this size.

This inefficiency can be overcome by approximating the exact equations with an evolution law for marginalized distributions. The simplest version of this is a system of equations for the probabilities of each node belonging to each group. The evolution law for these first-order statistics may be derived exactly, but the result involves second- and third-order statistics, i.e., the joint probabilities of group memberships for two and three nodes, respectively. From this, an approximate solution may be derived by specifying *closures* for the second- and third-order terms as functions of the first-order ones. However, although methods based on maintaining first-order statistics are efficient, they require external information to anchor nodes in various groups, or other *ad hoc* interventions, to prevent the distribution from converging to the uninformative state that all nodes are equally likely to be in any group. Such methods can work well in practice [7] when the data is sufficiently well understood, but any "probabilities" they produce are not meaningful.

This work focuses on the evolution law for the second-order statistics, the probabilities that pairs of nodes are in the same group. These quantities are of intrinsic interest, and Section 2 shows how they may also be used to form maximal-utility partitions of nodes into groups. Section 3 presents an approximation of these pairwise probabilities in the static case, while Section 4, being the bulk of the paper, addresses the dynamic case.

2 Expected utility of a group partition

When applying a group-finding algorithm to a graph G the usual output is a partition $\hat{\pi}$ of nodes into distinct groups. How does one assess the quality of such a

partition? Some group-finding algorithms are based on maximizing the *modularity* of the partition [2] with respect to *G*, so one could use this or a related statistic. When there is a ground-truth partition π available, it is preferable to compare $\hat{\pi}$ to π directly. A popular metric for this is the normalized mutual information [1]. More generally, one is free to choose a *utility function* for the computed partition $\hat{\pi}$ given the true partition π. Given such a utility function and a prior distribution on the true partition π, there is a simple optimality criterion for computed partitions $\hat{\pi}$: maximal *expected* utility. This is the criterion specified by Bayesian decision theory [8].

The utility of a computed partition depends on its intended use. A fairly generic use is this: for a given node *v*, one requires a classification of all the other nodes as being in the same group as *v* or not. To dramatize this, suppose all nodes in the true group of a randomly chosen node *v* are "bad" (and all other nodes "good"), whereas all nodes in the computed group of *v* are, after certain actions have been taken, "dead" (and all other nodes "alive"). When the computed partition $\hat{\pi}$ equals the true partition π all the bad nodes are dead and all the good ones are alive, which is optimal. Otherwise the utility depends on four parameters u_{BD}, u_{BA}, u_{GD}, and u_{GA} which specify the utility of a bad node being dead, a bad node being alive, etc. When *v* is equally likely to be any of the *n* nodes, the utility function may be expressed formally as

$$U^*\left(\hat{\pi}|\pi\right) \doteq \frac{1}{n}\sum_{v=1}^{n}\Big(u_{BD}\left|\hat{\pi}_v \cap \pi_v\right| + u_{BA}\left|\tilde{\hat{\pi}}_v \cap \pi_v\right| + u_{GD}\left|\hat{\pi}_v \cap \tilde{\pi}_v\right| + u_{GA}\left|\tilde{\hat{\pi}}_v \cap \tilde{\pi}_v\right|\Big), \quad (1)$$

where π_v is the set of nodes in the same true group as *v*, $\tilde{\pi}_v$ is the complement of π_v, etc., and $|C|$ is the number of elements of a set *C*. This utility function may be normalized and written as a sum over all groups $\hat{C} \in \hat{\pi}$ of this function:

$$U\left(\hat{C}|\pi\right) \doteq \frac{1}{2}\left(\sum_{C\in\pi}\left|\hat{C}\cap C\right|^2 - \theta\left|\hat{C}\right|^2 - (1-\theta)\left|\hat{C}\right|\right), \quad (2)$$

where

$$\theta = \frac{u_{GA}-u_{GD}}{\left(u_{GA}-u_{GD}\right)+\left(u_{BD}-u_{BA}\right)}. \quad (3)$$

The parameter θ encapsulates one's emphasis on killing bad nodes ($\theta = 0$) versus saving good ones ($\theta = 1$).

Now suppose that the ground truth partition π is unknown, but a graph *G* is given which provides information about π. In particular, one may use the probabilities $P(\pi|G)$ to compute the expected value of $U(\hat{C}|\pi)$ given *G*:

$$\mathbb{E}[U]\left(\hat{C}|G\right) = \sum_{\{v,w\}\subseteq\hat{C}}\left(p^{\{v,w\}} - \theta\right), \quad (4)$$

where $p^{\{v,w\}}$ is the probability that nodes *v* and *w* are in the same group of π, given *G*. Thus, the utility of a computed group $\hat{C}$ is proportional to the amount by which the average probability that two of its members are in the same group exceeds the threshold θ.

Equation (4) establishes that knowing $p^{\{v,w\}}$ for all node pairs $\{v,w\}$ provides a principled metric for assessing group partitions. Finding the best partition is then a problem of combinatorial optimization [2]. These pairwise probabilities $p^{\{v,w\}}$ are also of great interest themselves, and may be said to constitute a "soft" solution to group-finding problem. How, then, does one compute $p^{\{v,w\}}$? Section 3 provides an approximate answer for static graph data, and Section 4 for dynamic data.

3 Static case

Consider a scenario in which there are *n* nodes that are each assigned to one of *m* groups independently and with equal probability. For each pair of nodes the choice of whether a link exists is made independently, with link probability p_I if the nodes are in the same group, and p_O if they are in different groups. Typically $p_I > p_O$, so this is called the *homophily* model: a node "prefers" to be connected to nodes in the same group. This model will be denoted $\mathrm{H}(n,m,p_I,p_O)$. An instance of this model is a partition π of nodes into groups together with a graph *G*. For example, Figure 1 depicts an instance for the case of $n = 12$ nodes, $m = 3$ groups, $p_I = 0.8$, and $p_O = 0.1$, with groups indicated by node color.

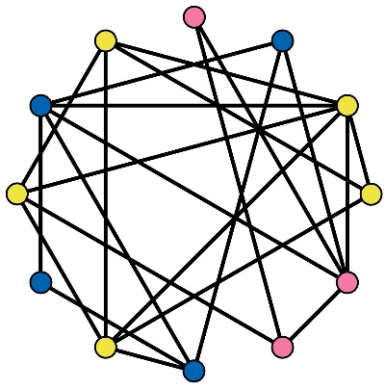

Figure 1: An instance of $\mathrm{H}(12,3,0.8,0.1)$

The key inference question is what can be determined about π when only *G* is given. One might ask, for example, for the probability p_i^v that the node *v* is in group *i* for *i* = 1 to *m*. By symmetry, however,

$p_i^v = 1/m$ for any v and i, regardless of G. More interesting are the second-order statistics p_{ij}^{vw} (the probability that node v is in group i and node w in group j). Symmetry reduces these statistics to functions of $p^{\{v,w\}}$, the probability that v and w are in the same group. As shown in Section 2, these pairwise probabilities enable group finding by providing the expected utility of a group partition.

There appears to be no simple formula for $p^{\{v,w\}}$ in terms of G. To obtain an exact value it is necessary to first compute the probabilities $\Pr(\pi|G)$ of each partition π given G, then sum this over all partitions with v and w in the same group to obtain $p^{\{v,w\}}$. This is a straightforward, but costly, computation: $\Pr(\pi|G)$ may be obtained by Bayesian inversion as follows,

$$\begin{aligned} &\Pr(G|\pi) = p_I^{e_I(G)}(1-p_I)^{e_I(\tilde{G})}\, p_O^{e_O(G)}(1-p_O)^{e_O(\tilde{G})}, \\ &\Pr(\pi) = \frac{(m)_{|\pi|}}{m^n}, \text{ and } \Pr(\pi|G) = \frac{\Pr(G|\pi)\Pr(\pi)}{\Pr(G)}. \end{aligned} \tag{5}$$

Here $e_I(G)$ is the number of links of G joining nodes in the same group (according to π), $e_I(\tilde{G})$ is the corresponding number of non-links, etc.; $|\pi|$ is the number of groups in π, and $(m)_r \doteq m!/(m-r)!$; and $\Pr(G)$ is a normalization constant.

To approximate $p^{\{v,w\}}$, let $p_M^{\{v,w\}}$ be the probability that v and w are in the same group given only that portion of G specified by the mask M. For any such mask,

$$p_M^{\{v,w\}} = \frac{\Lambda_M^{\{v,w\}}}{\Lambda_M^{\{v,w\}} + m - 1}, \tag{6}$$

where $\Lambda_M^{\{v,w\}}$ is the ratio of the likelihood of the evidence on $G \cap M$ under the hypothesis that v and w are in the same group to the likelihood under the hypothesis they are not. For example, let M = 2 denote the mask comprising the single link $\{v,w\}$, and $E(G)$ denote all links of G. Then

$$\Lambda_2^{\{v,w\}} = \begin{cases} \alpha \doteq p_I / p_O & \text{if } \{v,w\} \in E(G), \\ \beta \doteq (1-p_I)/(1-p_O) & \text{otherwise.} \end{cases} \tag{7}$$

Less trivially, let the mask M = 3 comprise all links adjacent to v and/or w. Of the $n-2$ nodes other than v and w, let n_0, n_1, and n_2 denote the number adjacent to neither, one, or both of v and w, respectively. Then the following, substituted into (6), produces the useful approximation $p_3^{\{v,w\}}$ to $p^{\{v,w\}}$:

$$\Lambda_3^{\{v,w\}} = \Lambda_2^{\{v,w\}} \left(\frac{\beta^2 + m - 1}{2\beta + m - 2}\right)^{n_0} \left(\frac{\alpha\beta + m - 1}{\alpha + \beta + m - 2}\right)^{n_1} \times \left(\frac{\alpha^2 + m - 1}{2\alpha + m - 2}\right)^{n_2}. \tag{8}$$

A more accurate, fourth-order correction may be derived as well, which is not based on a mask. It has 14 factors in place of the 3 in (8). Its use in group finding is the topic of a forthcoming paper.

4 Dynamic case

Now consider a case with n nodes and m groups, but allow the nodes to move from group to group, and the links to turn on and off. Let a be the rate of nodes moving: i.e., the probability of a node switching groups in an infinitesimal time dt is $a\,dt$. When a node switches, it is equally likely to join any of the other $m-1$ groups. Let λ_I be the rate at which links turn on for pairs of nodes in the same group, and μ_I be the rate at which they turn off. Let λ_O and μ_O be the corresponding rates for pairs of nodes in different groups. Finally, for any graph G, let

$$\gamma_X^{vw}(G) \doteq \begin{cases} \mu_X & \text{if } \{v,w\} \in E(G), \\ \lambda_X & \text{otherwise,} \end{cases} \tag{9}$$

for $X = I$ or O. This supplies the pertinent transition rate for any pair of nodes $\{v,w\}$.

This *dynamic homophily model* will be denoted $\mathcal{H}(n,m,a,\lambda_I,\mu_I,\lambda_O,\mu_O)$. Its instances are time histories of group partitions $\pi(t)$ and graphs $G(t)$. Figure 2 depicts an instance for the case of $n = 12$ nodes and $m = 3$ groups with rate parameters $a = 0.5$, $\lambda_I = 16$, $\mu_I = 4$, $\lambda_O = 2$, and $\mu_O = 18$ for t = 0 to 5.

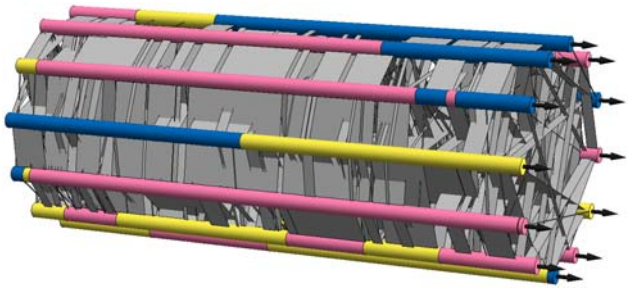

Figure 2: An instance of $\mathcal{H}(12,3,0.5,16,4,2,18)$

The key inference problem in this case is to determine information about $\pi(t)$ given the graph history up through some time t. Exact equations for performing this inference are given in [6] for a more general scenario (in which, e.g., links are of various types). Here they will

be expressed more succinctly. First, let ϕ be a function which maps every node to its group (capturing the information in π). Then define the following coefficients:

$$\alpha_{G,\phi\phi'} = \begin{cases} -na - \sum_{\{v,w\}\subseteq V} \gamma^{vw}_{\phi(v)\phi(w)} & \text{if } \phi = \phi', \\ a/(m-1) & \text{if } \phi = \phi' \text{ except at 1 node,} \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

and

$$\beta_{\phi,G^+G} = \begin{cases} \gamma^{vw}_{\phi(v)\phi(w)}(G) & \text{if } G^+ = G \text{ except at } \{v,w\}, \\ 0 & \text{otherwise,} \end{cases} \quad (11)$$

where $\gamma^{vw}_{ij} = \gamma^{vw}_I$ for $i = j$ and γ^{vw}_O otherwise.

Let $\tilde{p}^{\phi}(t)$ denote an unnormalized probability of the group assignment ϕ at time t, given the entire graph history up through time t. The evolution of $\tilde{p}^{\phi}(t)$ is governed by two equations. While $G(t)$ is constant, it obeys the differential equation

$$\dot{\tilde{p}}^{\phi} = \sum_{\phi'} \alpha_{G,\phi\phi'} \tilde{p}^{\phi'} . \quad (12)$$

When a link turns on or off, however, altering the graph from G to G^+, one applies

$$\tilde{p}^{\phi+} = \beta_{\phi,G^+G} \tilde{p}^{\phi} . \quad (13)$$

(For simultaneous, multiple link changes an average over all permutations of applying (13) is required.) These evolution equations are simple, but require a vast state space: there are m^n values of ϕ indexing $\tilde{p}^{\phi}$. Symmetry can reduce this somewhat, but to be practical, the state space must be much smaller still.

4.1 Exact second-order equations

Equations (12) and (13) may be summed over the groups of every node but one and then normalized to produce equations for the probability p_i^v that the node v is in group i for i = 1 to m. As in the static case, however, symmetry implies that $p_i^v(t) = 1/m$. Summing over the groups of every node but two yields equations for the second-order statistics p_{ij}^{vw}, which symmetry reduces to equations for the pairwise probabilities $p^{\{v,w\}}$. These equations involve the following auxiliary second-, third-, and fourth-order statistics:

$$\begin{aligned} q_{vw}^{vw} &= p^{\{v,w\}}\left(1 - p^{\{v,w\}}\right), \\ r_{vw}^{vx} &= p^{\{v,w,x\}} - p^{\{v,w\}} p^{\{v,x\}}, \text{ and} \\ s_{vw}^{xy} &= p^{\{v,w,x,y\}} + p^{\{v,w\}\{x,y\}} - p^{\{v,w\}} p^{\{x,y\}}. \end{aligned} \quad (14)$$

Here $p^{\{v,w\}\{x,y\}}$ denotes the probability that v and w are in one group, and x and y are together in a different one, etc.

While $G(t)$ is constant $p^{\{v,w\}}$ obeys the equation

$$\begin{aligned} \dot{p}^{\{v,w\}} = \frac{2am}{m-1}\left(\frac{1}{m} - p^{\{v,w\}}\right) - \gamma^{vw} q_{vw}^{vw} - \\ \sum_{x\neq v,w} \left(\gamma^{vx} r_{vw}^{vx} + \gamma^{wx} r_{vw}^{wx}\right) - \sum_{x,y\neq v,w} \gamma^{xy} s_{vw}^{xy} , \end{aligned} \quad (15)$$

and when a link turns on or off, one applies

$$p^{\{v,w\}+} = p^{\{v,w\}} + \begin{cases} \frac{\gamma^{vw} q_{vw}^{vw}}{\delta^{vw}} & \text{if } \{v,w\} \text{ flips,} \\ \frac{\gamma^{vx} r_{vw}^{vx}}{\delta^{vx}} & \text{if } \{v,x\} \text{ flips,} \\ \frac{\gamma^{xy} s_{vw}^{xy}}{\delta^{xy}} & \text{if } \{x,y\} \text{ flips.} \end{cases} \quad (16)$$

In these equations $\gamma^{vw} \doteq \gamma_I^{vw} - \gamma_O^{vw}$ (the dependence on G being understood), and δ^{vw} is the expected transition rate

$$\delta^{vw} \doteq \gamma_I^{vw} p^{\{v,w\}} + \gamma_O^{vw}\left(1 - p^{\{v,w\}}\right). \quad (17)$$

These marginalized equations are not closed: they depend on third- and fourth-order statistics which are not evolved as part of the system. One could solve the full equations (12) and (13) in order to get these higher-order statistics, and it has been checked that the result of evolving $p^{\{v,w\}}$ using (15) and (16) with these statistics is identical to marginalizing (12) and (13) directly. This verifies that the equations and their implementation are correct and exact, but to produce a practical method requires a *closure* of the system: i.e., a model for r_{vw}^{vx} and s_{vw}^{xy} in terms of $p^{\{v,w\}}$.

4.2 Second-order closure

The probability that v and w are in the same group and that x and y are in the same group is $p^{\{v,w,x,y\}} + p^{\{v,w\}\{x,y\}}$. If these events are treated as approximately independent, then this quantity may be approximated as $p^{\{v,w\}} p^{\{x,y\}}$: i.e., $s_{vw}^{xy} \approx 0$. Similarly one could argue that $r_{vw}^{vx} \approx 0$, though this is not as compelling. Making these approximations reduces the governing equations (15) and (16) to

$$\dot{p}^{\{v,w\}} = \frac{2am}{m-1}\left(\frac{1}{m} - p^{\{v,w\}}\right) - \gamma^{vw} p^{\{v,w\}}(1-p^{\{v,w\}}) \quad (18)$$

and

$$p^{\{v,w\}+} = \frac{\gamma_I^{vw} p^{\{v,w\}}}{\gamma_I^{vw} p^{\{v,w\}} + \gamma_O^{vw}\left(1-p^{\{v,w\}}\right)}. \quad (19)$$

Not only are these equations closed, they are local: the value of $p^{\{v,w\}}$ is determined only by the time history of the link between v and w. Thus (18) and (19) are the dynamic analog of (7). They provide a *prima facie*, baseline probability for v and w being in the same group.

Equation (18) can be solved exactly. Defining

$$k = \frac{m-1}{2am}\gamma^{vw}, \quad \text{and} \quad \Delta = \sqrt{(1+k)^2 - 4k/m}, \quad (20)$$

any initial value for $p^{\{v,w\}}$ on $[0,1]$ will converge to

$$p_c = \frac{2}{m(1+k+\Delta)} = \frac{1+k-\Delta}{2k}. \quad (21)$$

The explicit solution for $p^{\{v,w\}}(t)$ is

$$p^{\{v,w\}} = \begin{cases} 1 - \dfrac{2(m-1)/m}{1-k+\Delta\coth\left(\dfrac{am\Delta}{m-1}t\right)}, & p^{\{v,w\}} > p_c, \\ \dfrac{2/m}{1+k+\Delta\coth\left(\dfrac{am\Delta}{m-1}t\right)}, & p^{\{v,w\}} < p_c, \end{cases} \quad (22)$$

where t is initialized to make $p^{\{v,w\}}(t)$ agree with a specified initial condition.

4.3 Numerical results

In Section 3 the simple, local approximation $p_2^{\{v,w\}}$ was generalized by considering all links which share at least one node with $\{v,w\}$. The dynamic analog is to use $s_{vw}^{xy} \approx 0$ in (15) and (16), but to retain the r_{vw}^{vx} term. Before proceeding with this approximation, it is useful to examine the contributions of these terms, at least for an example case. Due to the expense of evolving (12) and (13) exactly, the example is limited to $n = 12$ nodes and $m = 3$ groups, which involves a state space with $3^{12} \approx 530{,}000$ components. The parameters for this example are the same as in depicted in Figure 2. A random instance of $\mathcal{H}(n,m,a,\lambda_I,\mu_I,\lambda_O,\mu_O)$ was generated in this case, and exact inference using (15) and (16) was performed. This inference was evolved to stationarity before results were collected. Figure 3 shows the histogram of values collected over time of the final summation in (15) for the case in which v and w are in the same group. A similar histogram was generated for the case in which they are in different groups, and it looks almost identical. In both cases the net effect of the fourth-order terms is both small and unbiased.

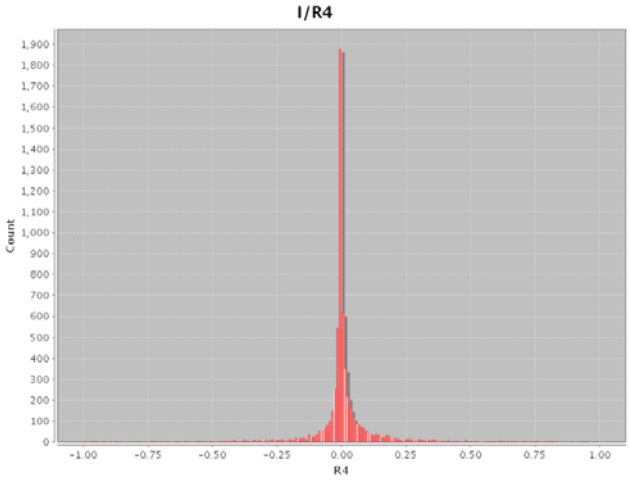

Figure 3: Fourth-order statistics within groups

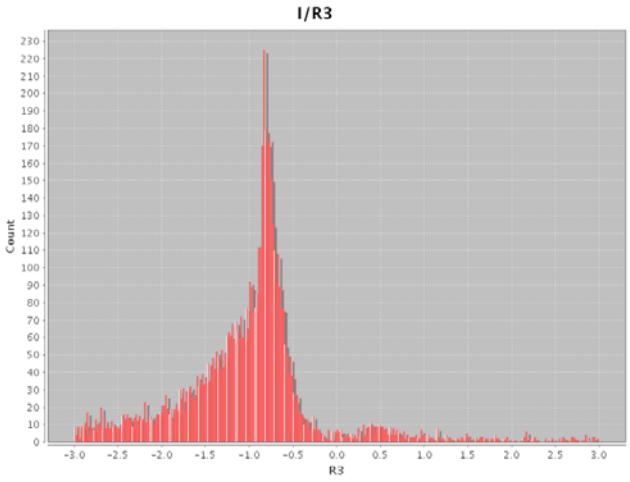

Figure 4: Third-order statistics within groups

A histogram of the second-to-last summation in (15) is shown in Figure 4 for the case in which v and w are in the same group. Figure 5 shows the corresponding histogram for different groups. These results indicate that the third-order effects are quite important: due to the negative sign preceding the second-to-last summation in (15), when v and w are in the same group, the predominantly negative values shown in Figure 4 drive $p^{\{v,w\}}$ higher, whereas the predominantly positive values in Figure 5 drive $p^{\{v,w\}}$ lower when they are in the different groups. These are just examples, but they corroborate the intuitive appeal of ignoring the influence of links which are connected to neither v nor w.

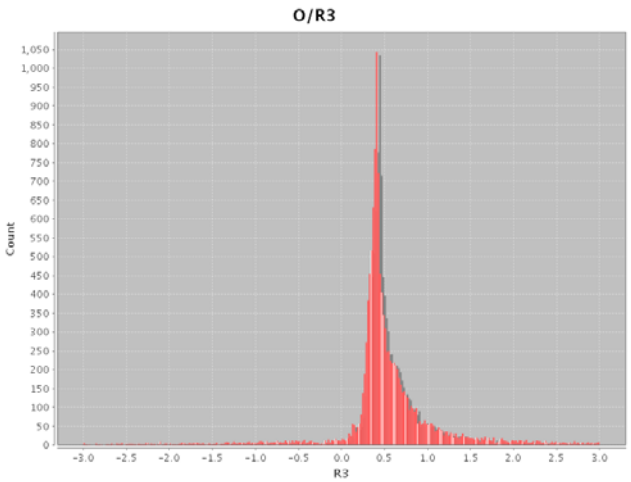

Figure 5: Third-order statistics between groups

4.4 Third-order closure: maximum entropy

There are five third-order statistics which, if the second-order statistics are given, must satisfy the following four equations

$$\begin{aligned} p^{\{v\}\{w\}\{x\}} + p^{\{w,x\}\{v\}} + p^{\{v,x\}\{w\}} + \\ p^{\{v,w\}\{x\}} + p^{\{v,w,x\}} = 1, \\ p^{\{w,x\}\{v\}} + p^{\{v,w,x\}} = p^{\{w,x\}}, \\ p^{\{v,x\}\{w\}} + p^{\{v,w,x\}} = p^{\{v,x\}}, \text{ and} \\ p^{\{v,w\}\{x\}} + p^{\{v,w,x\}} = p^{\{v,w\}}. \end{aligned} \tag{23}$$

This leaves one degree of freedom, represented by $p^{\{v,w,x\}}$. The constraint that the probabilities in (23) are non-negative imposes the following lower and upper bounds on $p^{\{v,w,x\}}$:

$$\begin{aligned} p^- &\doteq \frac{1}{2}\left(p^{\{w,x\}} + p^{\{v,x\}} + p^{\{v,w\}} - 1\right), \text{ and} \\ p^+ &\doteq \min\left(p^{\{v,w\}}, p^{\{v,x\}}, p^{\{w,x\}}\right). \end{aligned} \tag{24}$$

So a closure for $p^{\{v,w,x\}}$ should satisfy $p^- \le p^{\{v,w,x\}} \le p^+$. In particular, $p^- \le p^+$ is necessary for a consistent closure to exist. This condition is sufficient as well.

Assuming $p^- \le p^+$ what is the "best" value of $p^{\{v,w,x\}}$ to use? Many natural approximations (such as a symmetric version of $p^{\{v,w,x\}} \approx p^{\{v,w\}} p^{\{v,x\}}$) fail to satisfy $p^- \le p^{\{v,w,x\}} \le p^+$. The approximation $p^{\{v,w,x\}} \approx p^+$ has poor properties, as does $p^{\{v,w,x\}} \approx p_0^- \doteq \max(0, p^-)$. A least squares solution is possible, but inappropriate in this context (e.g., it does not guarantee positive solutions). A better principle to employ is maximum entropy [9].

When applying the maximum entropy principle, one needs a suitable underlying measure space. In this discrete case, this simply means a set of atomic events which are equally likely *a priori*. Such events arise naturally in this case: they have probabilities p_{ijk}^{vwx} which are each m^{-3} *a priori*. Thus the entropy is defined as

$$H = -\sum_{i,j,k} p_{ijk}^{vwx} \log p_{ijk}^{vwx}. \tag{25}$$

This may be re-written in terms of the five symmetric statistics occurring in (23), then reduced to a function of $p^{\{v,w,x\}}$. The derivative of this function is

$$\begin{aligned} \frac{dH}{dp^{\{v,w,x\}}} = \log\Bigg(\frac{(m-2)^2}{4(m-1)}\left(p^{\{w,x\}} - p^{\{v,w,x\}}\right) \times \\ \frac{\left(p^{\{v,x\}} - p^{\{v,w,x\}}\right)\left(p^{\{v,w\}} - p^{\{v,w,x\}}\right)}{p^{\{v,w,x\}}\left(p^{\{v,w,x\}} - p^-\right)^2}\Bigg). \end{aligned} \tag{26}$$

Provided $p_0^- < p^+$, the function in (26) is strictly decreasing from ∞ to $-\infty$ on $[p_0^-, p^+]$, so it has a unique zero within this interval, and this zero is where H attains its maximal value on $[p_0^-, p^+]$. Finding this zero requires solving a cubic equation.

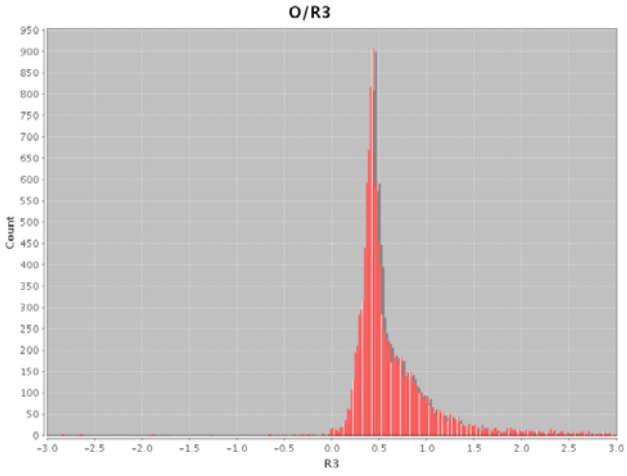

Figure 6: Third-order statistics between groups for maximum-entropy closure

How well does this closure approximate r_{vw}^{vx}? When exact values for $p^{\{v,w\}}$ are used in this maximum entropy closure to compute the summations plotted in Figure 4 and Figure 5 the result is surprisingly accurate. Figure 6 is the analog of Figure 5 for this closure, and shows good agreement: the results are predominantly positive, the peak is in the correct place, and the shape of the peak is correctly captured. A more rigorous test, however, would

be to not use exact values of $p^{\{v,w\}}$, but rather those obtained by actually evolving (15) and (16) using this closure. Unfortunately, this evolution of $p^{\{v,w\}}$ exhibits instabilities which cause the solution to blow up. Section 4.6 discusses why this occurs. There is another closure candidate, however, which can be evolved.

4.5 Third-order closure: extended state

Equations (15) and (16) involve third- and fourth-order statistics, which raises the question of what the evolution equations for these quantities are. These may be derived in a manner similar to second-order statistics. In particular, the third-order equations are

$$\begin{aligned}\dot{p}^{\{v,w,x\}} = {} & \frac{3am}{m-1}\left(\frac{p^{\{v,w\}}+p^{\{v,x\}}+p^{\{w,x\}}}{3m} - p^{\{v,w,x\}}\right) - \\ & \left(\gamma^{vw}q^{vw}_{vwx} + \gamma^{vx}q^{vx}_{vwx} + \gamma^{wx}q^{wx}_{vwx}\right) - \\ & \sum_{y\neq v,w,x}\left(\gamma^{vy}r^{vy}_{vwx} + \gamma^{wy}r^{wy}_{vwx} + \gamma^{xy}r^{xy}_{vwx}\right) - \sum_{y,z\neq v,w,x}\gamma^{yz}s^{yz}_{vwx},\end{aligned} \tag{27}$$

and

$$p^{\{v,w,x\}+} = p^{\{v,w,x\}} + \begin{cases} \dfrac{\gamma^{vw}q^{vw}_{vwx}}{\delta^{vw}} & \text{if } \{v,w\} \text{ flips,} \\ \dfrac{\gamma^{vy}r^{vy}_{vwx}}{\delta^{vy}} & \text{if } \{v,y\} \text{ flips,} \\ \dfrac{\gamma^{yz}s^{yz}_{vwx}}{\delta^{yz}} & \text{if } \{y,z\} \text{ flips,} \end{cases} \tag{28}$$

where

$$\begin{aligned} q^{vw}_{vwx} &= p^{\{v,w,x\}}\left(1-p^{\{v,w\}}\right), \\ r^{vy}_{vwx} &= p^{\{v,w,x,y\}} - p^{\{v,w,x\}}p^{\{v,y\}}, \text{ and} \\ s^{yz}_{vwx} &= p^{\{v,w,x,y,z\}} + p^{\{v,w,x\}\{y,z\}} - p^{\{v,w,x\}}p^{\{y,z\}}. \end{aligned} \tag{29}$$

Another approach to closure is to evolve both the $p^{\{v,w\}}$ and the $p^{\{v,w,x\}}$ statistics, cutting off the evolution equations for each at third order: i.e., $s^{xy}_{vw} \approx 0$, $r^{vy}_{vwx} \approx 0$, and $s^{yz}_{vwx} \approx 0$. Compared to the maximum entropy closure, this increases the space required from $O(n^2)$ to $O(n^3)$ in the dense case, but leaves the time required at $O(n^3)$.

This method exhibits less numerical instability than the maximum entropy method, and is at least qualitatively correct. For example, Figure 7 is the analog of Figure 5 for this closure. The peak is in the same place, and the values are still predominantly positive. However, the peak has diffused, particularly into the negative values. Figure 6 looks like a better approximation, but the comparison is unfair because the input values of $p^{\{v,w\}}$ were exact in Figure 6, but in Figure 7 are the result of an evolution which was not possible in the maximum entropy case.

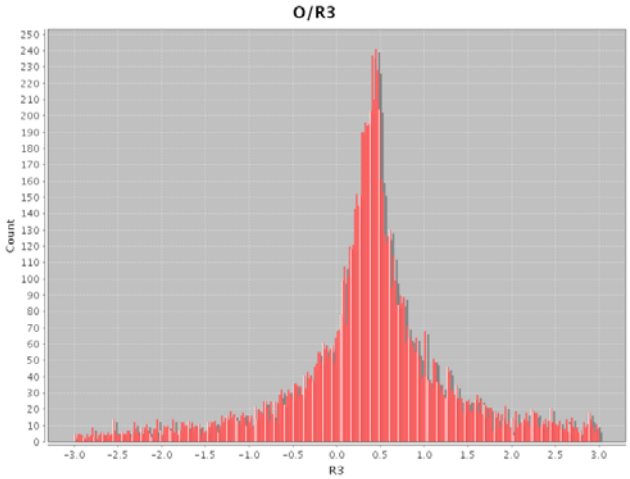

Figure 7: Third-order statistics between groups for extended-state closure

4.6 Consistency

The consistency condition $p^- \leq p^+$ mentioned in Section 4.4 may be re-written

$$\left|q^{\{u,v\}} - q^{\{u,w\}}\right| \leq q^{\{v,w\}} \leq q^{\{u,v\}} + q^{\{u,w\}}, \tag{30}$$

where $q^{\{v,w\}} \doteq 1 - p^{\{v,w\}}$. That is, the probabilities that nodes are in different groups satisfy the triangle inequality. In particular, if $p^{\{v,w\}} = 1$ then $p^{\{u,v\}} = p^{\{u,w\}}$ for every node $u \neq v,w$. This make sense: if v and w are definitely in the same group C, then both $p^{\{u,v\}}$ and $p^{\{u,w\}}$ mean "the probability that $u \in C$," so it would be illogical for these values to differ. Furthermore, in the special case $a = 0$, if $p^{\{v,w\}} = 1$ at some time, then it will remain 1, which implies $\dot{p}^{\{u,v\}} = \dot{p}^{\{u,w\}}$. The direct verification of this fact using (15) reduces to

$$\begin{aligned} \dot{p}^{\{u,v\}} - \dot{p}^{\{u,w\}} &= \overbrace{(T)}^{\text{third-order terms}} + \overbrace{(-T)}^{\text{fourth-order terms}} = 0, \\ \text{where } T &= \sum_{x\neq u,v,w}\left(\gamma^{wx} - \gamma^{vx}\right)r^{vx}_{uv}. \end{aligned} \tag{31}$$

In this case, the fourth-order terms play a crucial role in maintaining consistency, so it is not surprising that methods which truncate the fourth-order terms develop inconsistencies.

One way to maintain consistency is to return to the maximum entropy idea. The definition of H in (25) is a reasonable way to address the third-order closure problem, but the most natural definition of H uses atomic events with probabilities p^{ϕ}. This definition leads to a

closure which has a very simple form: the probability p^{π} of any partition π is given by

$$p^{\pi} = k \cdot (m)_{|\pi|} \prod_{\{v,w\} \prec \pi} \kappa^{\{v,w\}}, \qquad (32)$$

where k and each $\kappa^{\{v,w\}}$ for $\{v,w\} \subseteq V$ are positive, and $\{v,w\} \prec \pi$ means all $\{v,w\}$ which are in the same group in π. Although this seems promising, it is difficult to convert (32) into an efficient method.

5 Summary and future work

Network science is dominated by physicists, computer scientists, and sociologists, who each bring a certain set of techniques to the field. To the data fusion community, it is natural to frame group-finding as an inference problem: the graph data may be used to infer the (hidden) state of the group partition which influenced the graph's formation. This viewpoint is, in fact, present in the group-finding literature [10], but is somewhat anomalous. Where the techniques of the data fusion community seem particularly applicable, however, is the case of *tracking* groups in dynamically changing data. Aside from [6], this tracking approach does not appear in the literature.

Although [6] presents a formal method for tracking groups on dynamic networks, the technique is of theoretical interest only: the state space is too large for practical applications. The focus of this paper is suitable approximate techniques on a much smaller state space. Section 2 motivated the usefulness of the second-order statistic $p^{\{v,w\}}$, and Section 3 gave two methods for approximating it in the static case. Analogs of these methods were developed for the dynamic case in Section 4. The governing equations for $p^{\{v,w\}}$ require closures for third- and fourth-order statistics, and models are developed for the third-order statistics. The fourth-order statistics appear to be negligible, but play an important role in enforcing the logically necessary triangle inequality for the probabilities that nodes are in different groups.

The goal of this work is to develop a "Kalman filter for networks." In general, such terminology could be applied to any situation in which there is an evolving, hidden state of interest that influences dynamic network data, provided the solution has the form of a filter: i.e., a probability distribution over the hidden state which is evolved both when new data arrives and in between these arrivals. This paradigm has proven successful for traditional tracking problems, and seems promising for network problems too, whether this problem be tracking group membership, as studied here, or something else.

This research was supported by ONR Contract N0001409C0563.

References

[1] A. Lancichinetti and S. Fortunato, *Community detection algorithms: A comparative analysis*, Phys. Rev. E, Vol. 80, No. 5, 056117, 2009.

[2] S. Fortunato, *Community detection in graphs*, Phys. Rep., Vol. 486, No. 3-5, pp. 75-174, 2010.

[3] J. Hopcroft et al., *Tracking evolving communities in large linked networks*, Proc. Natl. Acad. Sci. U.S.A., Vol. 101, No. Suppl 1, pp. 5249-5253, 6 April 2004.

[4] G. Palla et al., *Quantifying social group evolution*, Nature, Vol. 446, No. 7136, pp. 664-667, 5 April 2007.

[5] A. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

[6] J. Ferry, *Group Tracking on Dynamic Networks*, Proc. 12th Int. Conf. on Information Fusion, Seattle, WA, July 6-9, 2009, pp. 930-937.

[7] G. Fenn et al., *Dynamic communities in multichannel data: An application to the foreign exchange market during the 2007–2008 credit crisis*, Chaos, Vol. 19, No. 3, 033119, 2009.

[8] J. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, Berlin, 1993.

[9] E. Jaynes, *Probability Theory: The Logic of Science (Vol 1)*, Cambridge University Press, Cambridge, 2003.

[10] M. Hastings, *Community detection as an inference problem*, Phys. Rev. E, Vol. 74, No. 3, 035102(R), 2006.