Discovering Hidden Groups in Communication Networks*

J. Baumes, M. Goldberg, M. Magdon-Ismail, and W. Wallace

Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180, USA.
Email: {baumej,goldberg,magdon}@cs.rpi.edu, wallaw@rpi.edu.

Abstract. We describe models and efficient algorithms for detecting groups (communities) functioning in communication networks which attempt to hide their functionality – *hidden groups*. Our results reveal the properties of the background network activity that make detection of the hidden group easy, as well as those that make it difficult.

1 Introduction.

The tragic events of September 11, 2001 underline the need for a tool which can be used for detecting groups that hide their existence and functionality within a large and complicated communication network, such as the Internet. In this paper, we view communication networks as random graphs. The nodes in this graph are the individuals or *actors* of the network, and an edge in the graph between two vertices represents a communication between the corresponding actors. We assume the communication infrastructure allows for any two actors communicate if they so choose. The randomness arises from the fact that two actors do not communicate in a deterministic fashion over time. Rather, at "random" times they are communicating, and at other times they are not. Thus, the graph that describes the communication dynamics of a network evolves with time according to some stochastic process. The question we ask is:

> *What properties of this evolving random communication graph will change if there is a hidden group attempting to camouflage its communications by embedding them into the background communications of the entire communication network?*

We must assume that any approach to detecting hidden groups should not rely on the semantic information contained in the communications. The reason is that communication within a hidden group on a public network is usually encrypted in some way, hence the semantic information may be either misleading or unavailable. The idea behind our approach is based upon the following observation:

> *Normal communications in the network are voluntary and "random;" however a hidden group communicates because it* **has to** *communicate (for planning or coordination).*

* This research was partially supported by NSF grants 0324947 and 0346341

Thus, the hidden group communication dynamics will display, out of necessity, certain non-random behavior. Detecting this non-random behavior will help us establish the presence of a hidden group, as well as identify its members. The property which we use to reveal non-randomness is the connectivity of certain subgraphs of the communication graph. Our analysis and simulations show that, for reasonable communication models of a society, it is possible to efficiently identify the hidden group. This forces a hidden group to face one of two outcomes, both of which are detrimental to the functioning of the hidden group: either continue with its planning or coordination (non-random communication dynamics) and risk being detected, or lower its planning or coordination activity to a level indistinguishable from the random background communication dynamics and risk not achieving its objectives. Our results indicate that there are three major factors that affect our ability to detect a hidden group.

i. The overall density of communications in the society. A higher density makes it more difficult to detect hidden groups. More specifically, there is a phase change at which the groups become significantly more difficult to detect with a relatively small increase in communication density.
ii. The presence of dense clusters. It is more difficult to detect a hidden group when the society communications are more structured into groups, keeping the overall communication density constant.
iii. The type of hidden group. We differentiate between *trusting* and *non-trusting* (or paranoid) groups. Trusting groups allow messages among group members to be delivered by non-group members, whereas non-trusting groups do not. Trusting groups tend to be benign, while non-trusting groups are more likely to be malicious. The surprising result is that it is **easier** to detect non-trusting groups; such groups are undermined by their own paranoia.

The implications of our results are two-fold. First, we can identify when it is feasible to detect hidden groups. Second, our results allow us to determine how long we must collect communication data to ensure that a hidden group is discovered.

The study of identifying hidden groups was initiated in [1] using Hidden Markov models. Here, our underlying methodology is based upon the theory of random graphs [2, 3]. We also incorporate some of the prevailing social science theories, such as homophily ([4]), by incorporating group structure into our model. A more comprehensive model of societal evolution can be found in [5, 6]. Other simulation work in the field of computational analysis of social and organizational systems [7–9] primarily deals with dynamic models for social network infrastructure, rather than the dynamics of the actual communication behavior. Our work is novel because we analyze the dynamics of communication intensities in order to detect hidden groups.

The outline of the remainder of the paper is as follows. First, we describe random graph models of communication networks. Then we discuss hidden groups and algorithms for detecting them, followed by extensive simulations that justify our conclusions in Section 5.

2 Random Graphs as Communication Models.

Social and information communication networks, such as the Internet and World Wide Web, are usually modeled by graphs (see [10, 7–9]), where the actors of the networks (people, IP-addresses, etc.) are represented by the vertices of the graph, and the connections between the actors are represented by the graph edges. Since we have no *a priori* knowledge regarding who communicates with whom, *i.e.* how the edges are distributed, it is appropriate to model the communications using a random graph. In this paper, we study hidden group detection in the context of two random graph models for the communication network. In describing these models, we will use standard graph theory terminology (see [11]), and its extension to *hypergraphs* (see [12]). In a hypergraph, the concept of an edge is generalized to a *hyperedge* which may join more than two vertices.

Random Model. A simple communication model is one where communications happen at random uniformly among all pairs of actors. Such a communication model can be represented by the random graph model developed and extensively studied by Erdős and Rényi, [13–15, 2]. In this model, the graph is generated by a random process in which an edge between every pair of nodes is generated independently with a given probability p. The class of graphs generated by such a random process is denoted $G(n, p)$.

Group Model. The $G(n, p)$ random graph model may not be a suitable model for large communication networks. Actors tend to communicate more often with certain actors and less frequently with others. In a more realistic model, actors will belong to one or more social groups where communication among group members is more frequent than communication among actors that do not belong to the same group. This leads us to the hypergraph model of the communication network, in which the actors associate themselves into groups. In this paper, we assume that each group is static and contains m actors. While this is a simplification, it serves to illustrate all the essential ideas and results without undue complication. A group of actors is represented by a hyperedge in the graph, and an actor may belong to zero or more hyperedges. The set of all hyperedges represents the structure of the communication network. Since groups tend to be small, it is appropriate to model the communications within a group as a $G(m, p_g)$, where p_g is the probability within the group. We also allow communication between two actors that do not share a group in common; we denote such communications as background. The probability of a background communication is p_b; we further assume that $p_b \ll p_g$ because intra-group communications are much more likely than extra-group communications.

Connectivity of Random Graphs. The key idea of our algorithms is based on the following observation. For any subset of actors in a random model network, it is very unlikely that this subset is connected during a "long" consecutive period of time cycles, while a hidden group must stay connected (for its operations) as long as it functions as a group. Thus, we summarize here some results from random graph theory that we will use regarding how the connectivity of a $G(n, p)$ depends on n and p, [13–15, 2]. These results are mostly asymptotic (with respect

to n) in nature, however, we use them as a guide that remains accurate even for moderately sized n.

Given a graph $G = \{V, E\}$, a subset $S \subseteq V$ of the vertices is connected if there exists a path in G between every pair of vertices in S. G can be partitioned into disjoint *connected components* such that every pair of vertices from the same connected component is connected and every pair of vertices in different connected components is not connected. The size of a component is the number of its vertices; the size of the largest connected component is denoted by $L(G)$.

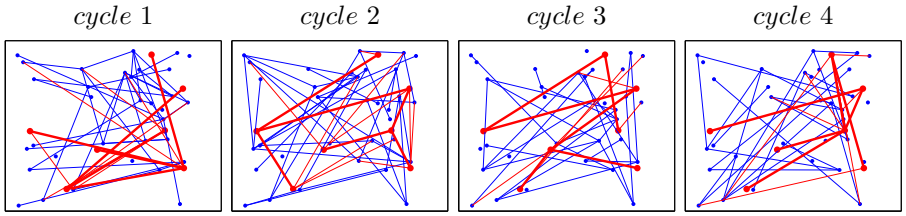
The remarkable discovery by Erdős and Rényi, usually termed *The Double Jump*, deals with the size of the largest component, and essentially states that $L(G)$ goes through two phase transitions as p increases beyond a critical threshold value. All the results hold asymptotically, with high probability, *i.e.*, with probability tending to 1 when $n \to \infty$:

$p = \frac{c}{n}$	$p = \frac{\ln n}{n} + \frac{x}{n}$, $x > 0$
$L(G(n,p)) = \begin{cases} O(\ln n) & 0 < c < 1 \\ O(n^{2/3}) & c = 1 \\ \beta(c)n & c > 1, \beta(c) < 1 \end{cases}$	$L(G(n,p)) = n$ with prob. $e^{-e^{-x}}$

Note that when $x \to \infty$, the graph is connected with probability 1. Thus, for $p =$ constant or $p = d \ln n/n$ with $d > 1$, the graph is asymptotically connected. However, when $p =$ constant, connectivity is exponentially more probable, which will have implications on our algorithms.

3 Hidden Groups.

The hidden group uses the normal society communications to camouflage communications of its members. On account of the planning activity, the hidden group members need to stay "connected" with each other during each "communication cycle." To illustrate the general idea, consider the following time evolution of a communication graph for a hypothetical society; here, communications among the hidden group are in bold, and each communication cycle graph represents the communications that took place during an entire time interval. We assume that information must be communicated among *all* hidden group members during one communication cycle.



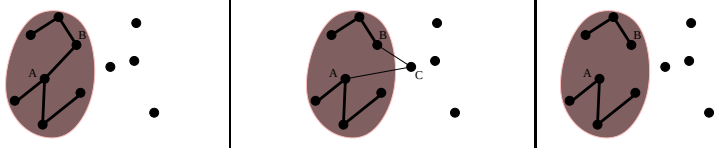
Note that the hidden group is connected in each communication cycle, *i.e.*, information can pass from any one group member to another, perhaps using other

actors as intermediaries, which is a requirement of our assumption that a message is passed from some group member to every other during every communication cycle. A hidden group may try to hide its existence by changing its connectivity pattern, or by throwing in "random" communications to non-hidden group members. For example, at some times the hidden group may be connected by a tree, and at other times by a cycle. None of these disguises changes the fact that the hidden group is connected, a property we will exploit in our algorithms.

Trusting vs. Non-Trusting Hidden Groups. Hidden group members may have to pass information to each other indirectly. Suppose that A needs to communicate with B. They may use a number of third parties to do this: $A \rightarrow C_1 \rightarrow \cdots \rightarrow C_k \rightarrow B$. *Trusting* hidden groups are distinguished from *non-trusting* ones by who the third parties C_i may be. In a trusting hidden group, the third parties used in a communication may be any actor in the society; thus, the hidden group members (A, B) trust some third-party couriers to deliver a message for them. In a non-trusting hidden group, *all* the third parties used to deliver a communication *must* themselves be members of the hidden group, *i.e.*, no one else is trusted. One expects that the more malicious a hidden group is, the more likely it is to be non-trusting.

Hidden groups that are non-trusting (vs. trusting) need to maintain a higher level of connectivity. We define three notions of connectivity as illustrated by the shaded groups in the following figure.

(a) Internally Connected (b) Externally Connected (c) Disconnected



A group is *internally connected* if a message may be passed between any two group members without the use of outside third parties. In the terminology of Graph Theory, this means that the subgraph induced by the group is connected. A group is *externally connected* if a message may be passed between any two group members, perhaps with the use of outside third parties. In Graph Theory terminology, this means that the group is a subset of a connected set of vertices in the communication graph. For example, in Figure (b) above, a message from A to B would have to use the outside third party C. A group is *disconnected* if it is not externally connected. The following observations are the basis for our algorithms for detecting hidden groups.

(i) A trusting hidden group is **externally connected** *in every communication cycle.*

(ii) A non-trusting hidden group is **internally connected** *in every communication cycle.*

We can now state the idea behind our algorithm for detecting a hidden group: a group of actors is *persistent* over communication cycles $1, \ldots, T$ if it is connected

in each of the communication graphs corresponding to each cycle. The two variations of the connectivity notion, internal or external, depend on whether we are looking for a non-trusting or trusting hidden group. Our algorithm is intended to identify persistent groups over a long enough time period as potential hidden groups. A hidden group can be hidden from view if, by chance, there are many other persistent subgroups in the society. In fact, it is likely that there will be many persistent subgroups in the society *during any given short time period*. However, these groups will be short-lived on on account of the randomness of the society communication graph. This is the reason our algorithm performs well over a long enough time period.

3.1 Detecting The Hidden Group.

The task of our algorithms is to efficiently identify maximal components that are persistent over a time period Π, and to to ensure with high probability that, over this time period, no persistent component can arise by chance, due to background communications.

Select Δ to be the smallest time-interval during which it is expected that information is passed among all group members. Index the communication cycles by $t = 1, 2, \ldots, T$. Thus, $T = \Pi/\Delta$. The communication data is represented by a series of communication graphs, G_t for $t = 1, 2, \ldots, T$. The vertex set for each communication graph is the set V of all actors. Below, we give algorithms to find persistent components.

(a) Externally persistent components

1: Ext_Persistent($\{G_t\}_{t=1}^{T}, U$)
2: //**Input:** Graphs $G_1, ..., G_T$ and $U \subseteq V$.
3: //**Output:** A partition of U.
4: Use DFS to get the connected components of every G_t;
5: Partition U into components $\{U_i\}_{i=1}^{K}$ such that two vertices are in the same component *iff* they are connected in every G_t;
6: **return** $\{U_i\}_{i=1}^{K}$;

(b) Internally persistent components

1: Int_Persistent($\{G_t\}_{t=1}^{T}, U$)
2: //**Input:** Graphs $G_1, ..., G_T$ and $U \subseteq V$.
3: //**Output:** A partition of U.
4: $\{U_i\}_{i=1}^{K}$ = Ext_Persistent($\{G_t\}_{t=1}^{T}, U$)
5: **if** $K = 1$, **then**
6: **return** $\{U_1\}$;
7: **else**
8: **return** $\cup_{k=1}^{K}$Int_Persistent($\{G_t\}_{t=1}^{T}, U_k$);
9: **end if**

The efficient implementation of step 5 in algorithm (a) above and the runtime analysis of these algorithms are postponed to a later exposition, as they are not central to our results.

Let h be the size of the hidden group we wish to detect. Let $X(t)$ denote the size of the largest persistent component over the communication cycles $1, \ldots, t$ that arises due to normal societal communications. $X(t)$ is a random variable with some probability distribution, since the communication graph of the society follows a random process. Given a confidence threshold, ϵ, we define the detection

time $\tau_\epsilon(h)$ as the time at which, with high probability governed by ϵ, the largest persistent component arising by chance in the background is smaller than h, *i.e.*,

$$\tau_\epsilon(h) = \min\{t : P[X(t) < h] \geq 1 - \epsilon\}.$$

Then, if after $\tau_\epsilon(h)$ cycles we observe a persistent component of size $\geq h$, we can claim, with a confidence $1 - \epsilon$, that this did not arise due to the normal functioning of the society, and hence must contain a hidden group. $\tau_\epsilon(h)$ indicates how long we have to wait in order to detect hidden groups of size h. Another useful function is $h_\epsilon(t)$, which is an upper bound for $X(t)$, with high probability, *i.e.*,

$$h_\epsilon(t) = \min\{h : P[X(t) < h] \geq 1 - \epsilon\}.$$

If, after a given time t, we observe a persistent component with size $\geq h_\epsilon(t)$, then with confidence at least $1 - \epsilon$, we can claim it to contain a hidden group. $h_\epsilon(t)$ indicates what sizes hidden group we can detect with only t cycles of observation. The previous approaches to detecting a hidden group assume that we know h or fix a time t at which to make a determination. By slightly modifying the definition of $h_\epsilon(t)$, we can get an even stronger hypothesis test for a hidden group. For any fixed $\delta > 0$, define

$$H_\epsilon(t) = \min\{h : P[X(t) < h] \geq 1 - \tfrac{\delta}{t^{1+\delta}}\epsilon\}.$$

Then one can show that if $X(t) \geq H_\epsilon(t)$ at any time, we have a hidden group with confidence $1 - \epsilon$.

Note that the computation of $\tau_\epsilon(h)$ and $h_\epsilon(t)$ constitute a pre-processing of the *society's communication* dynamics. This can be done either from a model (such as the random graph models we have described) or from the true, observed communications over some time period. More importantly, this can be done off-line. For a given realization of the society dynamics, let $T(h) = \min\{t : X(t) < h\}$. Some useful heuristics that aid in the computation of $\tau_\epsilon(h)$ and $h_\epsilon(t)$ by simulation can be obtained by assuming that $T(h)$ and $X(t)$ are approximately normally distributed, in which case,

Confidence level	$\tau_\epsilon(h)$	$h_\epsilon(t)$
50%	$E[T(h)]$	$E[X(t)]$
84.13%	$E[T(h)] + \sqrt{Var[T(h)]}$	$E[X(t)] + \sqrt{Var[X(t)]}$
97.72%	$E[T(h)] + 2\sqrt{Var[T(h)]}$	$E[X(t)] + 2\sqrt{Var[X(t)]}$

4 Experiments.

In our simulations, we fix the size n of the society at $n = 1000$. The results for both communication models, the random model and group model, are presented in parallel. For each model, multiple time series of graphs are generated for communication cycles $t = 1, 2, \ldots, T$, where $T = 100$. Depending on the nature of the plot, five to thirty time series were computed to smooth the plot reasonably. In order to estimate $h_\epsilon(t)$, we estimate $E[X(t)]$ by taking the sample average of

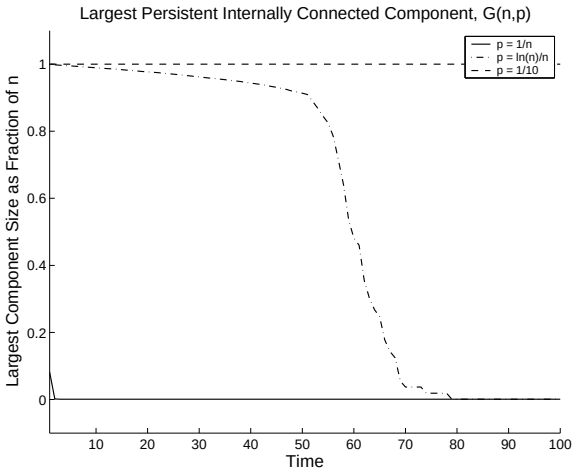
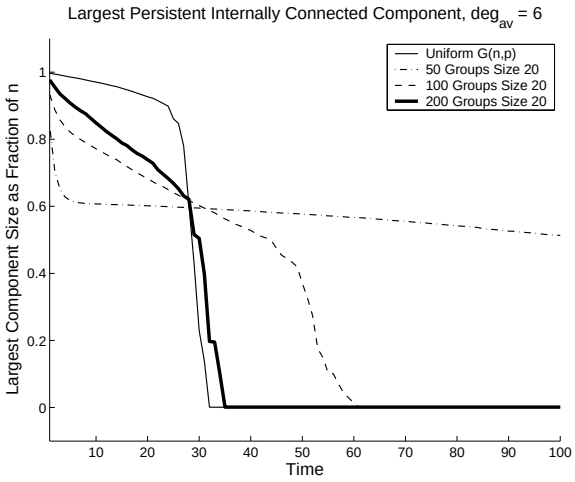


Fig. 1. The largest internally persistent component $E[X(t)]$ for the $G(n,p)$ model with $n = 1000$. The three lines represent $p = 1/n$, $p = \ln n/n$, and $p = 1/10$. Note the transition at $p = \log n/n$. When p is a constant, the graph is almost always connected.

the largest persistent component over communication cycles $1, \ldots, t$. Given h, the time at which the plot of $E[X(t)]$ drops below h indicates the time at which we can identify a hidden group of size $\geq h$.

We first describe the experiments with the random model ($G(n,p)$). The presence of persistantly connected components depends on the connectivity of the communication graphs over periods $1, 2, \ldots, T$. When the societal communication graph is connected for almost all cycles, we expect the society to generate many large persistent components. By the results of Erdős and Rényi described in Section 2, a phase transitions from short-lived to long-lived persistent components occur at $p = 1/n$ and $p = \ln n/n$. Accordingly, we present the results of the simulations with $p = 1/n$, $p = \ln n/n$, and $p = 1/10$ for $n = 1000$. The rate of decrease of $E[X(t)]$ is shown in Figure 1. For $p = 1/n$, we expect exponential or super-exponential decay in $E[X(t)]$ because $L(G)$ is at most a fraction of n. An abrupt phase transition occurs at $p = \ln n/n$. At this point the detection time begins to become large and unfeasible. For constant p, the graph is connected with probability equal to 1, and it becomes impossible to detect a hidden group using our approach without any additional information.

The parameters of the experiments with the group model are similar to that of the $G(n,p)$-model. We pick the group size m to be equal to 20. Each group is selected independently and uniformly from the entire set of actors; the groups may overlap; and each actor may be a member of zero or more groups. If two members are in the same group together, the probability that they communicate during a cycle is p_g, otherwise the probability equals p_b. It is intuitive that p_b is significantly bigger than p_b; we picked $p_b = 1/n$, so each actor has about one

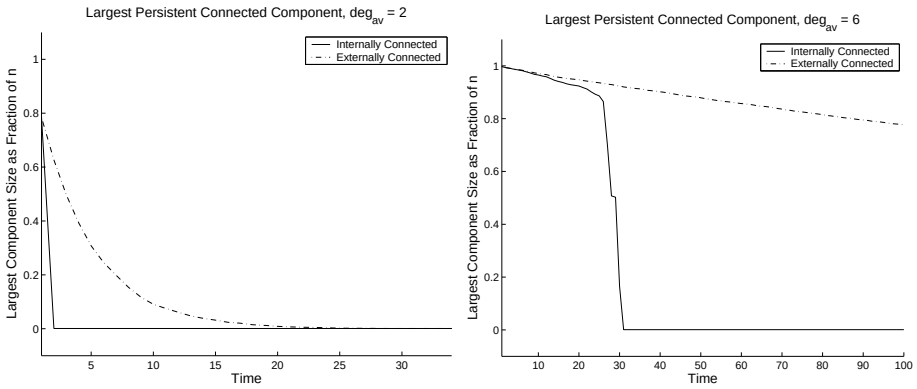


Groups	50	100	200	$G(n,p)$
$T(2)$	> 100	63	36	32

Fig. 2. Times of hidden group discovery for various amounts of group structure; each group is independently generated at random and has 20 actors. In all cases, $n = 1000$, $deg_{av} = 6$, and the group size $m = 20$. Note how, as the number of groups becomes large, the behavior tends toward the $G(n,p)$ case.

background communication per time cycle. The values of p_g that we use for the experiments are chosen to achieve a certain average number of communications per actor. In this way changes in the structure of the communication graph may be examined while keeping the density of communications constant. The average number of communications, or degree, per actor is set to six in the experiments. The number of groups is set to $g \in \{50, 100, 200\}$. These cases are compared to the $G(n,p)$ structure with an average of six commuications per actor. For the selected values of g, each actor is, on average, in 1, 2 and 4 groups, respectively. When g is 50, an actor is, on average, in approximately one group, and the overlaps of groups are small. However, when g is 200, each actor, on average, is in about 4 groups, so there is a significant amount of group overlap. The goal of experiments is to see the impact of g on finding hidden groups. Note that as as g increases, any given pair of actors tends to belong to at least one group together, so the communication graph tends toward a $G(n, p_g)$ graph.

We give a detailed comparison between the society with structure (group model) and the one without (random model) in Figure 2. The table shows $T(2)$, which is the time at which the size of the largest internally persistent component decreases to 1. This is the time at which any hidden group would be noticed, since the group would persist beyond the time expected in our model.



deg_{av}	$T(2)$ for trusting groups	$T(2)$ for non-trusting groups
2	28	2
6	> 100	32

Fig. 3. Times of hidden group discovery for non-trusting (internally connected) hidden groups and trusting (externally connected) hidden groups. In all cases the communication graphs are $G(n, p)$ with $n = 1000$.

We have also run the simulations for trusting groups. The results are shown in Figure 4. As the table shows, for the corresponding non-trusting communication model, the trusting group is much harder to detect.

Summary of Experimental Results. If the background society communications are dense, then it is harder to detect the hidden group. In fact, a phase transition occurs when the average node degree exceeds $\ln n$, at which point hidden groups of moderate size become hard to detect. If the hidden group is trusting, or the background society communications are structured, the hidden group is also harder to detect.

5 Discussion.

The experiments run in this study show that it is possible to discover malicious groups based upon structural properties of the communication graph, without using the contents of communications which may be misleading or difficult to decipher. Group identification done by an algorithm, such as the one proposed in this paper, could be used as the initial step in narrowing down the vast communication network to a smaller set of groups of actors. The communications among these potential hidden groups could then be scrutinized more closely.

Our results indicate a phase transition at $\ln n$ for the average number of communications that the average actor makes in one communication cycle. For moderately sized societies, from 10,000 to 100,000 actors, this phase transition occurs at about 10; *i.e.*, if the average number of communications is 10 per actor per communication cycle, then it becomes hard to detect the hidden group. Thus,

depending on the type of communication, the duration of the communication cycle may have to be made shorter to ensure that one is in the regime where a hidden group could be detected. Such an adjustment of the duration of a communicationb cycle may not match the time a hidden group ordinarily spends for a complete information exchange among its members. However closer to the planned event, they may need to communicate much more frequently, at which point they can be detected.

While the background communication density seems to be the dominant factor affecting our ability to detect the hidden group, we also find that if the society already has some structure (as with the group model), then it is harder to detect the hidden group. This result seems somewhat intuitive. However, a surprising result is that if the hidden group tries to hide all important communications within itself (a non-trusting group),it is *more easily* detected!

The value $T(h)$ that we compute in our simulations, is actually an upper bound on the time to hidden group discovery. We assume that the hidden group is clever enough to hide among the very "heart" of the communication network, the part that stays connected longest. If instead, the hidden group is extracted from the large component earlier, a simple extension of our algorithm would find the group much more quickly. Also note that this analysis uses no semantic information whatsoever. If there is any additional information available, such as certain individuals who should be watched, or certain type of messages, our algorithms can be modified to yiled a more efficient procedure for identification hidden groups.

Our results are of course only as accurate as our model is valid. However, we expect that the qualitative conclusions are robust with respect to the model, The extension of this work will explore more robust and realistic models of communication networks. Such models may explore the notion of a *conversation* between two actors, where their communication tends to persist over a certain length of time instead of being random at every time period. Also the groups can be made dynamic, where actors sometimes decide to enter or leave groups depending their preference.

The properties unique to the hidden group may also be modified for better results. A hidden group may not communicate at *every* time step, but may be connected more often than legitimate background groups. Also, a property not used in this analysis is that a hidden group is likely to be sparse (i.e. very nearly to a tree) to avoid detection. If there is a often-connected group that is sparse, it could be noted as more suspicious.

References

1. Magdon-Ismail, M., Goldberg, M., Wallace, W., Siebecker, D.: Locating hidden groups in communication networks using Hidden Markov Models. In: International Conference on Intelligence and Security Informatics (ISI 2003), Tucson, AZ (2003)
2. Bollobás, B.: Random Graphs, Second Edition. New york edn. Cambridge University Press (2001)

3. Janson, S., Luczak, T., Rucinski, A.: Random Graphs. Series in Discrete Mathematics and Optimization. Wiley, New york (2000)
4. Monge, P., Contractor, N.: Theories of Communication Networks. Oxford University Press (2002)
5. Goldberg, M., Horn, P., Magdon-Ismail, M., Riposo, J., Siebecker, D., Wallace, W., Yener, B.: Statistical modeling of social groups on communication networks. In: Inaugural conference of the North American Association for Computational Social and Organizational Science (NAACSOS 2003), Pittsburgh, PA (2003)
6. Siebecker, D.: A Hidden Markov Model for describing the statistical evolution of social groups over communication networks. Master's thesis, Rensselaer Polytechnic Institute, Troy, NY 12180 (2003) Advisor: Malik Magdon-Ismail.
7. Carley, K., Prietula, M., eds.: Computational Organization Theory. Lawrence Erlbaum associates, Hillsdale, NJ (2001)
8. Carley, K., Wallace, A.: Computational organization theory: A new perspective. In Gass, S., Harris, C., eds.: Encyclopedia of Operations Research and Management Science. Kluwer Academic Publishers, Norwell, MA (2001)
9. Sanil, A., Banks, D., Carley, K.: Models for evolving fixed node networks: Model fitting and model testing. Journal oF Mathematical Sociology **21** (1996) 173–196
10. Newman, M.E.J.: The structure and function of complex networks. SIAM Reviews **45** (2003) 167–256
11. West, D.B.: Introduction to Graph Theory. Prentice Hall, Upper Saddle River, NJ, U.S.A. (2001)
12. Berge, C.: Hypergraphs. North-Holland, New York (1978)
13. Erdős, P., Rényi, A.: On random graphs. Publ. Math. Debrecen **6** (1959) 290–297
14. Erdős, P., Rényi, A.: On the evolution of random graphs. Maguar Tud. Acad. Mat. Kutató Int. Kozël **5** (1960) 17–61
15. Erdős, P., Rényi, A.: On the strength of connectedness of a random graph. Acta Math. Acad. Sci. Hungar. **12** (1961) 261–267