Computing Subgraph Probability of Random Geometric Graphs with Applications in Quantitative Analysis of Ad Hoc Networks

Chang Wu Yu, *Member, IEEE*

***Abstract*—Random geometric graphs (RGG) contain vertices whose points are uniformly distributed in a given plane and an edge between two distinct nodes exists when their distance is less than a given positive value. RGGs are appropriate for modeling ad hoc networks consisting of n mobile devices that are independently and uniformly distributed randomly in an area. To the best of our knowledge, this work presents the first paradigm to compute the subgraph probability of RGGs in a systematical way. In contrast to previous asymptotic bounds or approximation, which always assume that the number of nodes in the network tends to infinity, the closed-form formulas we derived herein are fairly accurate and of practical value. Moreover, computing exact subgraph probability in RGGs is shown to be a useful tool for counting the number of induced subgraphs, which explores fairly accurate quantitative property on topology of ad hoc networks.**

***Index Terms*—Random geometric graphs, subgraph counting, subgraph probability, ad hoc networks, quantitative analysis.**

I. Introduction

A GEOMETRIC graph $G = (V, r)$ consists of nodes placed in a two-dimensional space R_2 and edge set $E = \{(i, j) \mid d(i, j) \leq r$, where $i, j \in V$ and $d(i, j)$ denotes the Euclidian distance between node i and node $j\}$. Let $X_n = \{x_1, x_2, \ldots, x_n\}$ be a set of independently and uniformly distributed random points. Here, $\Psi(X_n, r, A)$ is used to denote the random geometric graph (RGG) [16] of n nodes on X_n with radius r and placed in an area A. RGGs consider geometric graphs on random point configurations. Applications of RGGs include communications networks, classification, spatial statistics, epidemiology, astrophysics and neural networks [16].

A RGG $\Psi(X_n, r, A)$ is appropriate for modeling an ad hoc network $N = (n, r, A)$ consisting of n mobile devices with transmission radius r unit length that are independently and uniformly distributed randomly in an area A. When each vertex in $\Psi(X_n, r, A)$ represents a mobile device, each edge connecting two vertices represents a possible communication link as they are within the transmission range of each other. Fig. 1 displays a random geometric graph and its representing network. In the example, area A is a rectangle used to model the deployed area such as a meeting room. Area A, however, can be a circle, or any other shape, and even infinite space.

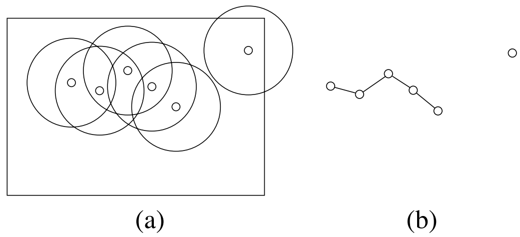


Fig. 1. (a) An ad hoc network $N = (6, r, A)$, where A is a rectangle. (b) Its associated random geometric graph $\Psi(X_6, r, A)$.

RGGs are different from conventional *random graphs* [2], [9], [15]. One random graph has two parameters n and p, where n denotes the number of nodes and p represents the probability of the existence of each possible edge. Edge occurrences in the random graph are independent of each other, unlike the case in RGGs. Therefore the fruitful results of random graphs cannot be directly applied to RGGs.

Counting the number of subgraphs in RGGs is of priority concern in quantitatively analyzing wireless ad hoc networks [12]–[14], [21], [22], [26], [27], [29], [31]. For example, the IEEE 802.11 CSMA/CA protocol suffers from the hidden and the exposed terminal problems [21], [22]. The hidden terminal problem is caused by concurrent transmissions of two nodes that cannot sense each other, but transmit to the same destination. Such two terminals are referred to here a *hidden-terminal pair*. Hidden-terminal pairs in an environment seriously results in garbled messages and increases communication delay, thus degrading system performance [12], [13], [22]. A hidden-terminal pair can be represented by a pair of edges $\{(x, y), (x, z)\}$ of $G = (V, E)$ such that $(x, y) \in E$ and $(x, z) \in E$, but $(y, z) \notin E$. In graph terms, such a pair of edges is an induced subgraph p_2 that is a path of length two (Fig. 2). Counting the occurrences of p_2 in a given RGG helps counting the number of hidden-terminal pairs in a network. The exposed terminal problem is due to prohibiting concurrent transmissions of two nodes that sense each other, but can transmit to different receivers without conflicts [21], resulting in unnecessary reduction in channel utilization and throughput. These nodes are referred to here as an *exposed-terminal set*. Similarly, the problem can be modeled as a subgraph M of $G = (V, E)$ with four vertices $\{x, y, z, w\} \subseteq V$ such that $\{(x, y), (y, z), (z, w)\} \subseteq E$, but $(x, z) \notin E$ and $(y, w) \notin E$ (Fig. 2).

Manuscript received 15 August 2008; revised January 31, 2009. This work was supported in part by Taiwan National Science Council under Contracts NSC 95-2221-E-216-021 and NSC 96-2221-E-216-009.

C. W. Yu is with the Department of Computer Science and Information Engineering, Chung Hua University, HsinChu 300, Taiwan, R.O.C. (e-mail: cwyu@chu.edu.tw).

Digital Object Identifier 10.1109/JSAC.2009.090904.

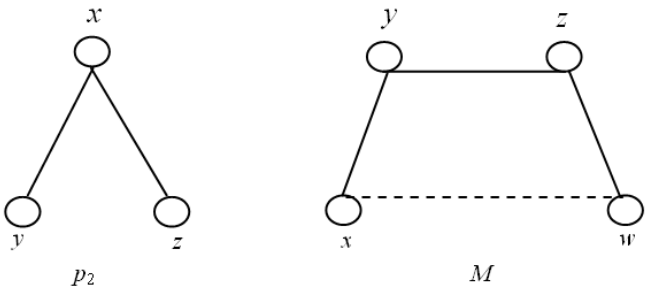


Fig. 2. Subgraphs of hidden-terminal pair p_2 and exposed-terminal set M.

The *subgraph probability* of a labeled subgraph $G = (V, E)$ in $\Psi(X_n, r, A)$ is defined formally as follows. Let $\Omega = \{G_1, G_2, \ldots, G_w\}$ represent every possible labeled simple graphs of $\Psi(X_n, r, A)$, where $X_n = \{x_1, x_2, \ldots, x_n\}$ and $w = 2^{\binom{n}{2}}$. For each labeled graph $G_k = (V_k, E_k)$ with $V_k = \{1, 2, \ldots, n\}$ in Ω, we have $E_k \subseteq V_k \times V_k$ and $d(x_i, x_j) \leq r$, where $(i, j) \in E_k$ and $1 \leq k \leq 2^{\binom{n}{2}}$. Given a subgraph $G_x = (V_x, E_x)$ where $V_x \subseteq \{1, 2, \ldots, n\}$ and $E_x \subseteq V_x \times V_x$, the subgraph probability of G_x in $\Psi(X_n, r, A)$, denoted by $\Pr(G_x)$, is summing up the probabilities of all label graphs in Ω whose induced subgraphs by V_x are identical to G_x. Specifically, we have $\Pr(G_x) = \sum_{\forall G \in \Omega \text{ and } G_{V_x} = G_x} \Pr(G)$.

Counting the number of subgraphs of RGGs has received considerable attention [16]. Penrose demonstrated that, for arbitrary feasible connected subgraph Γ with k vertices, the number of induced subgraphs isomorphic to Γ satisfies a Poisson limit theorem and a normal limit theorem [16]. To our knowledge, results of previous studies are all asymptotic or approximate without accurate closed-form functions.

This work presents the first paradigm for exact computing the subgraph probability of RGGs in a systematical way. With the derived results, counting the numbers of some specific induced subgraphs in RGGs and their applications in quantitative analysis on topology of ad hoc networks are discussed. In contrast to previous asymptotic bounds or approximation, which always assume that the number of nodes in the network tends to infinity, the closed-form formulas derived here are fairly accurate and of practical value. Moreover, by applying the proposed paradigm, a great deal of topologies of ad hoc networks can be analyzed.

Our definition of random geometric graphs $\Psi(X_n, r, A)$ is different from those of Poisson point process [1], [8], which assume that the distribution of n points (vertices) on a possibly infinite plane follows a Poisson distribution with parameter (the given density). In Poisson point process, the number of vertices can only be a random number rather than a tunable parameter. In practice, however, some wireless network modeling requires a fixed input n (for example, when the number of deployed sensors is fixed) or a limited deployed area (for example, when the deployed area is a 100m$\times$100m square). Torus convention models the network topology so that nodes near the border are considered as close to nodes at the opposite border; they are allowed to establish links as well. When the deployed area is large, the border effects don't really matter. Consequently, torus convention is adopted here to remove border effects such that the deployed area appears to be homogeneous at any point [1], [11]. Note that part of the early draft of this work can be found in [28], [30].

The rest of the paper is organized as follows. Section II introduces definitions and notations. Section III then briefly surveys pertinent literature on RGGs. Next, Section IV presents a novel paradigm for computing the subgraph probability of RGGs. Based on the derived results, in Section V, the number of some given induced subgraphs in RGGs are counted along with their applications in ad hoc networks discussed as well. Conclusions are finally drawn in Section VI along with areas for future research.

II. Definitions and Notations

Definitions and notations used in this work are defined formally here. A *graph* $G = (V, E)$ consists of a finite nonempty vertex set V and edge set E of unordered pairs of distinct vertices of V. A graph is *simple* if it has no loops and no two of its links join the same pair of vertices. A graph $G = (V, E)$ is *labeled* when the $|V|$ vertices are distinguished from one another by names such as $v_1, v_2, \ldots, v_{|V|}$. Two labeled graphs $G = (V_G, E_G)$ and $H = (V_H, E_H)$ are *identical*, denoted by $G = H$ if $V_G = V_H$ and $E_G = E_H$. A graph $H = (V_H, E_H)$ is a *subgraph* of $G = (V_G, E_G)$ if $V_H \subseteq V_G$ and $E_H \subseteq E_G$. The *size* of any set S is denoted by $|S|$. The *degree* of a vertex v in graph G is the number of edges incident with v. A *leaf* is a vertex of degree one.

The notation $\binom{n}{m}$ denotes the number of ways to select m from n distinct objects. An edge e of G is said to be *contracted* if it is deleted and its ends are identified. A subgraph of G is said to be *contracted* if all its edges are contracted successively in any order. Two subgraphs are *disjoint* if their edge sets are disjoint. The subgraph of $G = (V, E)$ whose vertex set is $V' \subseteq V$ and whose edge set is the set of those edges of G that have both ends in V' is called the subgraph of G induced by V', denoted by $G_{V'}$. An induced subgraph that is a path of length i is denoted by p_i. Similarly, an induced subgraph that is a cycle of length i is denoted by c_i; c_3 is often called a *triangle*. A set of vertices is *independent* if no two of them are adjacent. An induced subgraph which is an independent set of size i is denoted by I_i. A *complete graph* is a simple graph whose vertices are pair-wise adjacent; the unlabeled complete graph with n vertices is denoted K_n. A *forest* is an acyclic graph. A *tree* is a connected acyclic graph.

In a graph $G = (V, E)$, a set $S \subset V$ is a *dominating set* if every vertex not in S has a neighbor in S. The *domination number* $\gamma(G)$ is the minimum size of a dominating set in G. Let $G = (V_G, E_G)$ and $H = (V_H, E_H)$ be graphs. Let f be a 1-1 mapping of V_G onto V_H, and g a 1-1 mapping of E_G onto E_H. Let θ denote the ordered pair (f, g). We say that θ is an *isomorphism* of G onto H if the following condition holds: the vertex x is incident with the edge e in G if and only if the vertex $f(x)$ is incident with the edge $g(e)$ in H. The notational conventions used in the paper can be found in [3].

III. Related Work in RGGs

We summary related results as follows. A book and several papers written by Penrose [16]–[19] provide and explain the theory of random geometric graphs. Graph problems considered in the book include subgraph and component counts, vertex degrees, cliques and colorings, minimum degree, the largest component, partitioning problems, and connectivity and the number of components.

For n points uniformly randomly distributed on a unit cube in $d \geq 2$ dimensions, Penrose [19] showed that the resulting geometric random graph G is k-connected and G has minimum degree k at the same time when $n \to \infty$. In [5], [6], Díaz *et al*. discussed many layout problems including minimum linear arrangement, cutwidth, sum cut, vertex separation, edge bisection, and vertex bisection in random geometric graphs. In [7], Díaz *et al.* considered the clique or chromatic number of random geometric graphs and their connectivity.

Some results of RGGs can be applied to the connectivity problem of ad hoc networks. In [20], Santi and Blough discussed the connectivity problem of random geometric graphs $\Psi(X_n, r, A)$, where A is a d-dimensional region with the same length size. In [1], Bettstetter investigated two fundamental characteristics of wireless networks: its minimum node degree and its k-connectivity. In [8], Dousse *et al*. obtained analytical expressions of the probability of connectivity in the one dimension case. In [10], Gupta and Kumar have shown that if $r = \sqrt{\frac{\log n + c(n)}{\pi n}}$, then the resulting network is connected with high probability if and only if $c(n) \to \infty$. In [24], Xue and Kumar have shown that each node should be connected to $\Theta(\log n)$ nearest neighbors in order that the overall network is connected.

Recently, Yen and Yu have analyzed link probability, expected node degree, and expected coverage of MANETs [26]. In [25], Yang has obtained the limits of the number of subgraphs of a specified type which appear in a random graph.

IV. Computing Subgraph Probability of RGGs

In the section, a novel paradigm for exactly computing subgraph probability of RGGs is proposed. After discussing the dependence and independence of edge probabilities, we deal with subgraphs with three vertices. Then the proposed paradigm is introduced. For simplicity, we always assume that A is sufficiently large to properly contain a circle with radius r in a $\Psi(X_n, r, A)$ throughout the paper (that implies $|A| \geq \pi r^2$).

A. Dependence and independence of edge probabilities

Hereafter, notation E_i (E_i') denotes the event of the occurrence (absence) of edge e_i in a RGG. Since we adopt torus convention to avoid border effects, the probability of a single-edge (u, v) in a RGG is equivalent to the probability of that v is located in the circle centered at u with radius r. The result is summarized in Theorem 1.

Theorem 1: We have $\Pr(E_j) = \pi r^2/|A|$, for an arbitrary edge $e_j = (u, v)$ and $u \neq v$, in a $\Psi(X_n, r, A)$.

The following theorem shows that the occurrences of arbitrary pair-wise edges in RGGs are independent even if they share one end vertex.

Theorem 2 ([26]): For arbitrary two distinct edges $e_i = (u, v)$ and $e_j = (w, x)$ in a $\Psi(X_n, r, A)$, we have $\Pr(E_iE_j) = \Pr(E_i)\Pr(E_j)$.

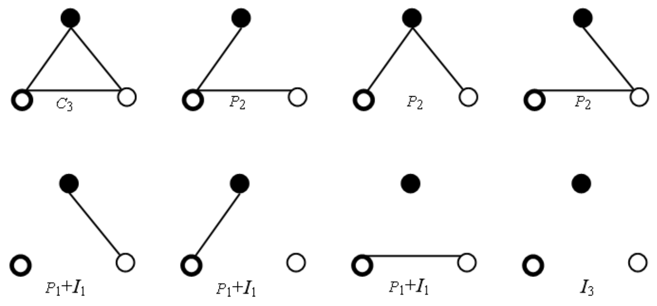


Fig. 3. Eight labeled subgraphs with three vertices.

Combining Theorem 1 and 2, we obtain the probability of two-edge subgraphs immediately.

Corollary 3: For arbitrary two distinct edges $e_i = (u, v)$ and $e_j = (w, x)$ in a $\Psi(X_n, r, A)$, we have $\Pr(E_iE_j) = (\pi r^2/|A|)^2$.

Theorem 2 indicates that the occurrences of two distinct edges in RGGs are independent. At first glance, the obtained result is somewhat difficult to be accepted as facts for some scholars. In [20], Santi and Blough claimed that the occurrences of two distinct edges $e_1 = (u, v)$ and $e_2 = (u, w)$ are correlated and dependent by observing that "if $d(u, v) < d(u, w)$, then the existence of e_2 implies the existence of e_1." The above statement is true, but it does not necessarily imply that the statement: "the existence of e_2 implies the existence of e_1" is also true. Their arguments are somewhat misleading.

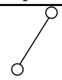
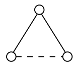
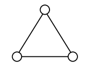
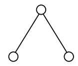
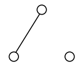
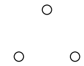
Although Corollary 3 has assured the independence of arbitrary two edges in RGGs, the occurrences of three or more edges in RGGs are dependent (which will be shown in the next subsection). Also note that, when border effects are considered, Yen and Yu have shown the dependence of two-edge probability in RGGs [26].

B. Compute the probabilities of three-vertex subgraphs in RGGs

In the subsection, the probabilities of eight labeled subgraphs with three vertices are computed, and will form a basis for computing the probability of larger subgraphs (with more than three vertices) in the proposed paradigm. Based on the number of edges included, the subgraphs can be classified into four groups: (1) a triangle c_3, (2) a single edge with an isolated vertex $p_1 + I_1$, (3) an induced path of length two p_2, and (4) three isolated vertices I_3 (Fig. 3). In this work, it is possible for graphs that are not identical to have essentially the same diagram, with the exception that their vertices have different labels. Sometimes, instead of using labels, vertices in the following figures are shaded differently to distinguish them. The probabilities of these eight subgraphs in a $\Psi(X_n, r, A)$ can be computed by manipulating elementary geometric techniques (described in Theorems 6-10) and Table I summaries the results beforehand.

To obtain the probability of c_3, we need Lemmas 4 and 5. Here, if one of two equal-sized circles in the place contains the center of the other, we call them two properly intersecting circles.

TABLE I
PROBABILITIES OF LABELED SUBGRAPHS WITH THREE OR FEWER VERTICES IN A RGG.

Notation	p_1	E^2	c_3	p_2	$p_1 + I_1$	I_3
G						
$\Pr(G)$	$\frac{\pi r^2}{\lvert A\rvert}$	$\left(\frac{\pi r^2}{\lvert A\rvert}\right)^2$	$\frac{\left(\pi-\frac{3\sqrt{3}}{4}\right)\pi r^4}{\lvert A\rvert^2}$	$\frac{\left(\frac{3\sqrt{3}}{4}\right)\pi r^4}{\lvert A\rvert^2}$	$\frac{\pi^2}{\lvert A\rvert}\left(1-\frac{\pi^2}{\lvert A\rvert}-\frac{\frac{3\sqrt{3}}{4}r^2}{\lvert A\rvert}\right)$	$1-\frac{\pi r^4}{\lvert A\rvert}-\frac{\frac{3\sqrt{3}}{4}}{\lvert A\rvert^2}\pi r^4$

Lemma 4 ([31]): The expected overlapped area of two properly intersecting circles with radius r is $\left(\pi - \frac{3\sqrt{3}}{4}\right) r^2$.

Lemma 5: For three distinct edges $e_i = (u,v)$, $e_j = (u,w)$, and $e_k = (v,w)$ in a $\Psi(X_n, r, A)$, we have $\Pr(E_iE_j \mid E_k) = \left(\pi - \frac{3\sqrt{3}}{4}\right) r^2/|A|$, where $u \neq v \neq w$.

Proof: Let O (O') be the circle centered at v (w) with radius r. Since $\Pr(E_iE_j \mid E_k)$ is the probability of that u is located in the overlapped area of two properly intersecting circles O and O', we have $\Pr(E_iE_j \mid E_k) = \left(\pi - \frac{3\sqrt{3}}{4}\right) r^2/|A|$ by Lemma 4. ■

The probability of the first labeled subgraph c_3 can then be obtained.

Theorem 6: For three distinct edges $e_i = (u,v)$, $e_j = (u,w)$, and $e_k = (v,w)$ in a $\Psi(X_n, r, A)$, we have $\Pr(E_iE_jE_k) = \left(\pi - \frac{3\sqrt{3}}{4}\right)\pi r^4/|A|^2$, where $u \neq v \neq w$.

Proof: By Theorem 1 and Lemma 5, we have $\Pr(E_iE_jE_k) = \Pr(E_k)\Pr(E_iE_j \mid E_k) = (\pi r^2/|A|) \times \left(\left(\pi - \frac{3\sqrt{3}}{4}\right) r^2/|A|\right) = \left(\pi - \frac{3\sqrt{3}}{4}\right)\pi r^4/|A|^2$. ■

Next, we derive the probability for a single edge with an isolated vertex ($p_1 + I_1$).

Theorem 7: For three distinct edges $e_i = (u,v)$, $e_j = (u,w)$, and $e_k = (v,w)$ in a $\Psi(X_n, r, A)$, we have $\Pr(E_iE'_jE'_k) = \frac{\pi r^2}{|A|}\left(1 - \frac{\pi r^2}{|A|} - \frac{3\sqrt{3}r^2}{4|A|}\right)$, where $u \neq v \neq w$.

Proof: Let A_u (A_v) denote the area of the circle O (O') centered at u (v) with radius r. Let $A_u \cap A_v$ ($A_u \cup A_v$) denote the overlapped area (union area) of these two circles. Since the expected size of $A_u \cap A_v$ is $\left(\pi - \frac{3\sqrt{3}}{4} r^2\right)$ (by Lemma 4) and $|A_u \cap A_v| + |A_u \cup A_v| = 2\pi r^2$, the expected size of $A_u \cup A_v$ is $2\pi r^2 - \left(\pi - \frac{3\sqrt{3}}{4}\right) r^2 = \left(\pi + \frac{3\sqrt{3}}{4}\right) r^2$. Accordingly, we have $(E_iE'_jE'_k) = \Pr(E_i) \times \Pr(E'_jE'_k \mid E_i) = \Pr(E_i) \times (1 - \Pr(w \text{ is located in } A_u \cup A_v)) = \Pr(E_i) \times \left(1 - \left(\pi + \frac{3\sqrt{3}}{4}\right) r^2/|A|\right) = (\pi r^2/|A|) \times \left(1 - \left(\pi + \frac{3\sqrt{3}}{4}\right) r^2/|A|\right) = \frac{\pi r^2}{|A|}\left(1 - \frac{\pi r^2}{|A|} - \frac{3\sqrt{3}r^2}{4|A|}\right)$. ■

The next subgraph considered is an induced path p_2, which can be used to model a hidden-terminal pair.

Theorem 8: For arbitrary three distinct edges $e_i = (u,v)$, $e_j = (u,w)$, and $e_k = (v,w)$ in a $\Psi(X_n, r, A)$, we have $\Pr(E_iE_jE'_k) = \left(\frac{3\sqrt{3}}{4}\right)\pi r^4/|A|^2$, where $u \neq v \neq w$.

Proof: Since $\Pr(E_iE_jE'_k) = \Pr(E_iE_j) - \Pr(E_iE_jE_k) = \Pr(E_i)Pr(E_j) - \Pr(E_iE_jE_k)$ by Theorem 2, we have $\Pr(E_iE_jE'_k) = (\pi r^2/|A|)^2 - \left(\pi - \frac{3\sqrt{3}}{4}\right)\pi r^4/|A|^2 = \left(\frac{3\sqrt{3}}{4}\right)\pi r^4/|A|^2$ by Corollary 3 and Theorem 6. ■

Finally, the last subgraph probability of three isolated vertices I_3 is computed.

Theorem 9: For arbitrary three distinct edges $e_i = (u,v)$, $e_j = (u,w)$, and $e_k = (v,w)$ in a $\Psi(X_n, r, A)$, we have $\Pr(E'_iE'_jE'_k) = 1 - \frac{\pi r^4}{|A|} - \frac{\frac{3\sqrt{3}}{4}}{|A|^2}\pi r^4$, where $u \neq v \neq w$.

Proof: There are $2^3 = 8$ possible subgraphs and four groups with any three vertices (Fig. 3). Accordingly, we have $1 = \Pr(E_iE_jE_k) + \binom{3}{1} \times \Pr(E_iE_jE'_k) + \binom{3}{1} \times \Pr(E_iE'_jE'_k) + \Pr(E'_iE'_jE'_k)$. By applying Theorems 6-8, we obtain $\Pr(E'_iE'_jE'_k) = 1 - \Pr(E_iE_jE_k) - 3 \times Pr(E_iE_jE'_k) - 3 \times \Pr(E_iE'_jE'_k) = 1 - \left(\pi - \frac{3\sqrt{3}}{4}\right)\pi r^4/|A|^2 - 3 \times \left(\frac{3\sqrt{3}}{4}\right)\pi r^4/|A|^2 - 3 \times \frac{\pi r^2}{|A|}\left(1 - \frac{\pi r^2}{|A|} - \frac{3\sqrt{3}r^2}{4|A|}\right) = 1 - \frac{\pi r^4}{|A|} - \frac{\frac{3\sqrt{3}}{4}}{|A|^2}\pi r^4$. ■

In Subsection IV-A, we have obtained that the occurrences of two distinct edges in a $\Psi(X_n, r, A)$ are independent (Theorem 2). The next theorem, however, shows that edge independence does not exist for subgraphs with arbitrary three or more edges.

Theorem 10: The occurrences of arbitrary three distinct edges in a $\Psi(X_n, r, A)$ are dependent.

Proof: It is sufficient to show that $\Pr(E_iE_jE_k) \neq \Pr(E_i)\Pr(E_j)\Pr(E_k)$ for arbitrary three distinct edges. Suppose that the three edges in a $\Psi(X_n, r, A)$ are e_i, e_j, and e_k. Four disjoint cases are possible: (C_1) These edges form a cycle; (C_2) Every pair in these edges shares no end point; (C_3) Only two of three edges share one end point; (C_4) These edges form a tree or a path.

Case C_1: These edges form a cycle. Since $\Pr(E_iE_jE_kC_1) = \Pr(E_iE_jE_k \mid C_1) \times \Pr(C_1) = \Pr(C_1)$ and $\Pr(E_i)\Pr(E_j)\Pr(E_k)\Pr(C_1) = (\pi r^2/|A|)^3 \times \Pr(C_1)$ by Theorem 1, we have $\Pr(E_iE_jE_kC_1) \neq \Pr(E_i)\Pr(E_j)\Pr(E_k)\Pr(C_1)$.

Case C_2: These edges share no end point. Obviously, we have $\Pr(E_iE_jE_k \mid C_2) = \Pr(E_i)\Pr(E_j)\Pr(E_k) = (\pi r^2/|A|)^3$ (by Theorem 1) and $\Pr(E_iE_jE_kC_2) = \Pr(E_iE_jE_k \mid C_2) \times \Pr(C_2) = (\pi r^2/|A|)^3 \times \Pr(C_2)$.

Case C_3: Only two of three edges share one end point. Without loss of generality, let $e_i = (u,v)$ and $e_j = (u,w)$ and $e_k = (x,y)$, where e_i and e_j share vertex u. Since $G_{\{(u,v),(u,w)\}}$ and $G_{\{(x,y)\}}$ are two disjoint subgraphs, we have $\Pr(E_iE_jE_k \mid C_3) = \Pr(E_iE_j)\Pr(E_k)$ and $\Pr(E_iE_jE_kC_3) = \Pr(C_3)\Pr(E_iE_jE_k \mid C_3) = \Pr(C_3)\Pr(E_iE_j)\Pr(E_k) = \Pr(E_i)\Pr(E_j)\Pr(E_k)\Pr(C_3)$ by Theorems 1 and 2.

Case C_4: These edges form a tree or a path. (1) If it is a tree, let $e_i = (u,v)$ and $e_j = (w,v)$ and $e_k = (x,v)$, where e_i, e_j, and e_k share vertex v; (2) If it is a path, let $e_i = (u,v)$

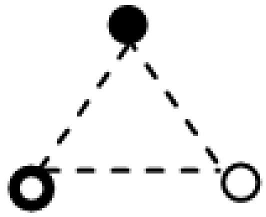


Fig. 4. The complete class graph representing the set of eight subgraphs in Fig. 3.

and $e_j = (v, w)$ and $e_k = (w, x)$, where e_i and e_j share vertex v and e_j and e_k share vertex w. For both cases, we can easily derive that $\Pr(E_iE_jE_k \mid C_4) = \Pr(E_i)\Pr(E_j)\Pr(E_k)$ and $\Pr(E_iE_jE_kC_4) = \Pr(E_i)\Pr(E_j)\Pr(E_k)\Pr(C_4)$.

On the contrary, suppose that these three edges are independent; we have $\Pr(E_iE_jE_k) = \Pr(E_i)Pr(E_j)Pr(E_k)$. According to above discussion, $\Pr(E_iE_jE_k) = \Pr(E_iE_jE_kC_1) + \Pr(E_iE_jE_kC_2) + \Pr(E_iE_jE_kC_3) + \Pr(E_iE_jE_kC_4) = \Pr(E_iE_jE_kC_1) + \Pr(E_i)\Pr(E_j)\Pr(E_k)(\Pr(C_2) + \Pr(C_3) + \Pr(C_4)) = \Pr(E_iE_jE_kC_1) + \Pr(E_i)\Pr(E_j)\Pr(E_k)(1 - \Pr(C_1)) = \Pr(E_i)\Pr(E_j)\Pr(E_k)$. Consequently, we have $\Pr(E_iE_jE_kC_1) = \Pr(E_i)\Pr(E_j)\Pr(E_k)\Pr(C_1)$, which is a contradiction. Finally, we conclude that $\Pr(E_iE_jE_k) \neq \Pr(E_i)\Pr(E_j)\Pr(E_k)$. ■

C. Paradigm for computing subgraph probability

In this subsection, a paradigm for computing subgraph probability is presented. First, a graph drawing convention, which is helpful for describing the proposed paradigm, is given. A *solid line* denotes an edge of G; a *broken line* denotes a possible edge between them; two vertices without a line denote a *non-edge* of G. A *class graph* $G = (V, E_S, E_B)$ consists of a vertex set V and two disjoint edge sets E_S and E_B, where E_S (E_B) denotes a set of solid-line edges (broken-line edges) joining two vertices of V. A *complete class graph* is a class graph whose vertices are pair-wise adjacent with either a solid line or a broken line. Any class graph can be modeled by the graph drawing convention. For example, the complete class graph shown in Fig. 4 denotes the set of eight labeled subgraphs depicted in Fig. 3.

Some operators and notation of class graphs are defined here. The *union* of two class graphs G_a and G_b, denoted $G_a + G_b$, is the set whose elements are exactly the graphs in either G_a or G_b. The difference of two class graphs G_a and G_b, denoted $G_a - G_b$, is the set containing exactly those elements in G_a that are not in G_b. When G is a class graph, $\Pr(G)$ denotes the probability of the occurrence of $G_x \in G$ in $\Psi(X_n, r, A)$. If every element in G_a is also in G_b, we have $G_a \subseteq G_b$. Evidently, if $G_a \subseteq G_b$ then $\Pr(G_a) \leq \Pr(G_b)$, and if G_a is isomorphic to G_b then $\Pr(G_a) = \Pr(G_b)$. The union and difference of class graphs can be represented by the graph drawing convention. For example, the equation $\Pr(E_2) = \Pr(c_3) + \Pr(p_2)$ is valid (Table I), which can be represented in two different forms: the first-type and the second-type graph derivations (Fig. 5).

In fact, this first-type (second-type) graph derivation can be applied on any broken-line edge (non-edge) of any class graph. Specifically, for any $e \in E_B$, we have $G(V, E_S, E_B) = G_a(V, E_S \cup \{e\}, E_B - \{e\}) + G_b(V, E_S, E_B - \{e\})$. Or for any $e \notin E_S \cup E_B$, we have $G(V, E_S, E_B) = G_a(V, E_S, E_B \cup \{e\}) - G_b(V, E_S \cup \{e\}, E_B)$ equivalently.

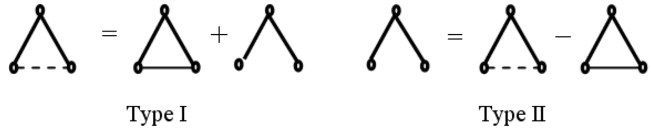


Fig. 5. Two graph derivations for class graphs.

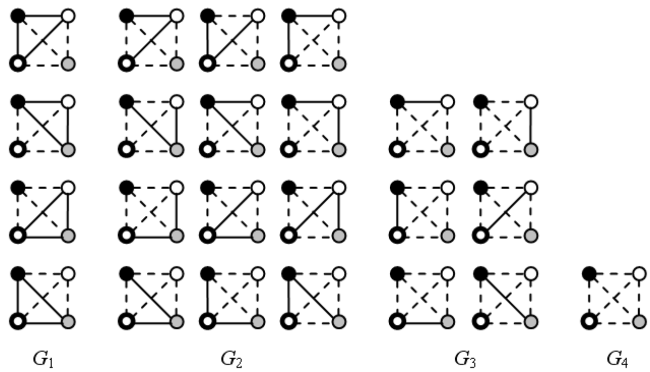


Fig. 6. Four equivalence sets in Group I.

1) The paradigm:

Given a subgraph $G = (V, E)$, the paradigm computes its probability $\Pr(G)$ of a RGG by exploiting the following three steps:

(1) *Preprocessing step*: First, generate all complete class graph set CG_n with the same labeled vertex set where $n = |V|$. Find out all the equivalence sets such that the underline graph of each class graph is isomorphic to each other in same set. Next, compute the subgraph probability $\Pr(x)$ for each equivalence set x. Moreover, one element from each distinct equivalence set is selected to form a *basis* $\{G_1, G_2, \ldots, G_k\}$ of CG_n.

(2) *Decomposing step*: Decompose G into a linear combination of the selected basis of CG_n: $c_1G_1 + c_2G_2 + \cdots + c_kG_k$ by recursively applying the second-type graph derivations for each non-edge in class graph G.

(3) *Manipulating step*: Compute $\Pr(G) = c_1\Pr(G_1) + c_2\Pr(G_2) + \cdots + c_k\Pr(G_k)$, where $\Pr(G_i)$ is obtained in the preprocessing step, for $1 \leq i \leq k$.

2) Applying the paradigm for computing four-vertex subgraph probability:

In this subsection, we apply the proposed paradigm and compute the probability of subgraph M (which has been used to model an exposed-terminal set in Fig. 2) as an example.

a) Preprocessing step: The complete class graph set CG_4 including $2^6 = 64$ class graphs which can be further categorized into three groups:

Group I: The complete class graphs in CG_4 consist of at least one vertex incident with three-broken lines (Fig. 6).

Totally, there are four equivalence sets in Group I (Fig. 6). Note that the probabilities of these four class graphs are equal to one of three or fewer vertex subgraphs in Table I; specifically, for arbitrary four distinct nodes $S = \{u, v, w, x\}$

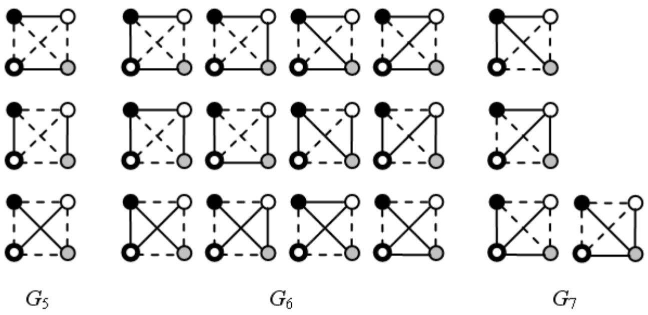


Fig. 7. Three equivalent sets in Group II.

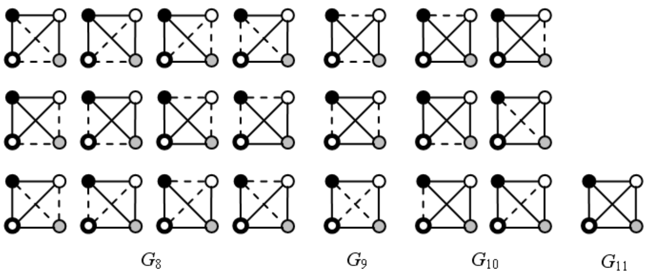


Fig. 8. Four equivalent sets in Group III.

in a $\Psi(X_n, r, A)$, we have $\Pr(G_S$ is a specific element in $G_1) = \Pr(c_3)$, $\Pr(G_S$ is a specific element in $G_2) = \Pr(E_2)$, $\Pr(G_S$ is a specific element in $G_3) = \Pr(p_1)$, and $\Pr(G_S$ is a specific element in $G_4) = 1$.

Group II: Every vertex in one of the complete class graphs in CG_4 is incident with at least one solid line and any subgraph induced by solid lines is a forest (without cycles). There are three equivalence sets in Group II (Fig. 7). Their probabilities can be obtained by using the results shown in Subsections IV-A and IV-B. Specifically, for arbitrary four distinct nodes $S = \{u, v, w, x\}$ in a $\Psi(X_n, r, A)$, we have $\Pr(G_S$ is a specific element in $G_5) = (\pi r^2/|A|)^2$ by Theorems 1 and 2. Also, we have $\Pr(G_S$ is a specific element in $G_6) = \Pr(G_S$ is a specific element in $G_7) = (\pi r^2/|A|)^3$.

Group III: Every vertex in one of the complete class graphs in CG_4 is incident with at least one solid line and the subgraph induced by its solid lines contains a cycle.

There are four equivalence sets in Group III (Fig. 8). Their probabilities can be calculated by Theorems 11-15.

Theorem 11: For arbitrary four distinct nodes u, v, w, and x in a $\Psi(X_n, r, A)$, we have $\Pr(G_S$ is a specific element in $G_8) = \left(\pi - \frac{3\sqrt{3}}{4}\right)\frac{\pi^2 r^6}{|A|^3}$, where $S = \{u, v, w, x\}$.

Proof: Since any specific element in G_8 consists of a triangle and a solid edge; its probability can be obtained by applying Theorem 6 and Theorem 1 once; that is, $\Pr(G_S$ is a specific element in $G_8) = \left(\pi - \frac{3\sqrt{3}}{4}\right)\frac{\pi r^4}{|A|^2} \times \frac{\pi r^2}{|A|} = \left(\pi - \frac{3\sqrt{3}}{4}\right)\frac{\pi^2 r^6}{|A|^3}$. ■

Theorem 12: For arbitrary four distinct nodes u, v, w, and x in a $\Psi(X_n, r, A)$, we have $\Pr(G_S$ is a specific element in $G_{10}) = \left(\pi - \frac{3\sqrt{3}}{4}\right)^2 \times \frac{\pi r^6}{|A|^3}$, where $S = \{u, v, w, x\}$.

Proof: Since any specific element in G_{10} consists of two triangles with a broken line; we can obtain the probability by applying Theorem 1 once and Lemma 5 twice; that is, $\Pr(G_S$ is a specific element in $G_{10}) = \left(\left(\pi - \frac{3\sqrt{3}}{4}\right)\frac{r^2}{|A|}\right)^2 \times \frac{\pi r^2}{|A|} = \left(\pi - \frac{3\sqrt{3}}{4}\right)^2 \times \frac{\pi r^6}{|A|^3}$. ■

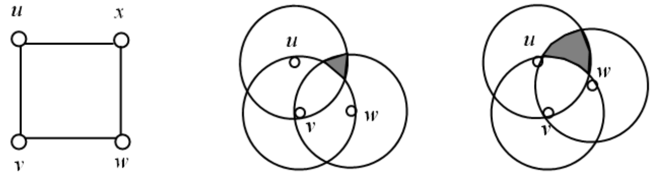


Fig. 9. (a) A cycle of length four. (b) The circle model of its subgraph including u, v, and w. (c) The maximum size of the gray area.

Theorem 13: For arbitrary four distinct nodes u, v, w, and x in a $\Psi(X_n, r, A)$, we have $\Pr(G_S = c_4) \leq \frac{\sqrt{3}}{8}\frac{\pi^2 r^6}{|A|^3}$, where $S = \{u, v, w, x\}$.

Proof: Suppose that nodes u, v, w, and x form a cycle of length four c_4 (Fig. 9(a)). These four nodes need to be placed properly near to each other in order to form the cycle. However, nodes u and w need to keep a distance away from each other for avoiding an edge occurred between them (Fig. 9(a)). Also, nodes v and x have to meet the same requirement.

The subgraph induced by three nodes $\{u, v, w\}$, which form an induced path of length two p_2, is considered first. With respect to the circle model of the subgraph, the remaining node x should be placed in the gray area of Fig. 9(b) so that node x will connect to u and w, but excluding v. Consequently, we have $\Pr(G_S = c_4) \leq \Pr(G_{S-\{x\}} = p_2) \times \Pr($the remaining vertex x is in the gray area). To obtain a bound of $\Pr(G_S = c_4)$, we need to estimate the size of the gray area.

Consider the circle model again. To maximize the size of the gray area, node w (u) should be placed as close to the circle centered at u (w) as possible but still in the coverage of the circle centered at v. Similarly, node v should be placed as far away from both of u and w as possible but still in the intersection area of the two circles centered at u and w, respectively. As a result, when the centers of nodes u, v, and w are placed in the intersection points of the three circles depicted in Fig. 9(c), the maximum size of the gray area is achieved. Specifically, the maximum size of the gray area is $\pi r^2/6$, which can be obtained by subtracting the size of the intersection area of three circles from the size of the lens of two circles in Fig. 9(c) [34].

At last, we have $\Pr(G_S = c_4) \leq \Pr(G_{S-\{x\}} = p_2) \times \Pr($the remaining vertex x is in the gray area $) \leq \frac{3\sqrt{3}}{4}\frac{\pi r^4}{|A|^2} \times \frac{\pi r^2}{6|A|} = \frac{\sqrt{3}}{8}\frac{\pi^2 r^6}{|A|^3}$ by Table I. ■

Theorem 14: For arbitrary four distinct nodes u, v, w, and x in a $\Psi(X_n, r, A)$, we have $\Pr(G_S)$ is a specific element in $G_{11}) \leq \left(\frac{\pi^2}{2} - \frac{7\sqrt{3}\pi}{8} + \frac{9}{8}\right)\frac{\pi r^6}{|A|^3}$, where $S = \{u, v, w, x\}$.

Proof: The four nodes $\{u, v, w, x\}$ need to be placed sufficiently near to each other in order to form a K_4 (Fig. 10(a)). First, the three nodes $\{u, v, w\}$ must be a triangle c_3. The circle model for c_3 can be presented by intersections of three equal circles (Fig. 10(b)).

Since the subgraph K_4 consists of a triangle c_3 and another nearby vertex x, we have $\Pr(GS$ is a specific element in $G_{11}) \leq \Pr(G_{S-\{x\}} = c_3) \times \Pr(x$ is near c_3 sufficiently). Note that $\Pr(x$ is near c_3 sufficiently) is the probability of

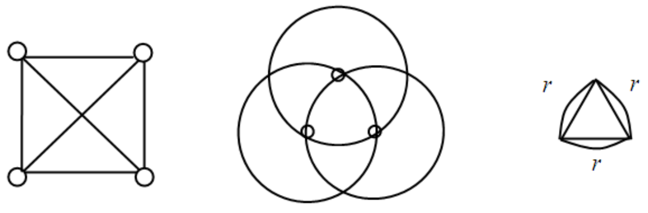


Fig. 10. (a) K_4. (b) Its circle model. (c) Reuleaux triangle.

Fig. 11. Representing G_9 by a combination of c_4, G_{10}, and G_{11}.

putting the center of x in the common intersection of three equal circles; and the largest area of the common intersection, called Reuleaux triangle [34], is $\left(\frac{\pi-\sqrt{3}}{2}\right)r^2$ (it is easily obtained by summing up the area of a equilateral triangle and three areas of a circular segment with opening angle $\pi/3$). Therefore, we have $\Pr(x \text{ is near } c_3 \text{ sufficiently}) \leq \left(\frac{\pi-\sqrt{3}}{2}\right)r^2/|A|$. In a sequel, we have $\Pr(G_S \text{ is a specific element in } G_{11}) \leq \left(\pi - \frac{3\sqrt{3}}{4}\right)\pi r^4/|A|^2 \times \left(\frac{\pi-\sqrt{3}}{2}\right)r^2/|A| = \left(\frac{\pi^2}{2} - \frac{7\sqrt{3}\pi}{8} + \frac{9}{8}\right)\frac{\pi r^6}{|A|^3}$ by Theorem 6. ■

Finally, we deal with the last equivalence set G_9 and the result is given below.

Theorem 15: For arbitrary four distinct nodes u, v, w, and x in a $\Psi(X_n, r, A)$, we have $\Pr(G_S \text{ is a specific element in } G_9) \leq \left(2\pi^2 - \frac{23\sqrt{3}\pi}{8} + \frac{27}{8}\right)\frac{\pi r^6}{|A|^3}$, where $S = \{u, v, w, x\}$.

Proof: Since each element in G_9 represents a combination of elements in c_4 and G_{10} and G_{11} (Fig. 11), we have $\Pr(G_S \text{ is a specific element in } G_9) = \Pr(G_S = c_4) + 2 \times \Pr(G_S \text{ is a specific element in } G_{10}) - \Pr(G_S \text{ is a specific element in } G_{11}) \leq \left(\frac{\sqrt{3}}{8}\right)\frac{\pi^2 r^6}{|A|^3} + 2 \times \left(\pi - \frac{3\sqrt{3}}{4}\right)^2 \times \frac{\pi r^6}{|A|^3} = \left(2\pi^2 - \frac{23\sqrt{3}\pi}{8} + \frac{27}{8}\right)\frac{\pi r^6}{|A|^3}$ by Theorems 12 and 13. ■

To sum up, the subgraph probabilities of nine of eleven complete class graphs: G_1, G_2, …, G_{11}, which form a basis of CG_4, can be formulated by closed forms. For the remaining two complete class graphs (e.g., G_9 and G_{11}), upper bounds are obtained instead.

b) Decomposing step:

Given any four-vertex class graph G, the step decomposes it by iteratively applying the second-type graph derivation for each non-edge in G and the subsequently generated class graphs until the resulting graphs are all complete class graphs. Since $\{G_1, G_2, \ldots, G_{11}\}$ is a basis of CG_4 according to the preprocessing step, a linear combination for $G = c_1G_1 + c_2G_2 + \cdots + c_{11}G_{11}$ are obtained.

The following example (Fig. 12) shows how the proposed paradigm computes the probability of subgraph M successfully (the exposed-terminal set in Fig. 2). In fact, the probability of any subgraph (or class graph) with four vertices can be calculated in the similar way.

c) Manipulating step:

Accordingly, we have the desired subgraph probability $\Pr(M) = \Pr(G_6) - 2 \times \Pr(G_8) + \Pr(G_{10}) = \left(\pi r^2/|A|\right)^3 -$

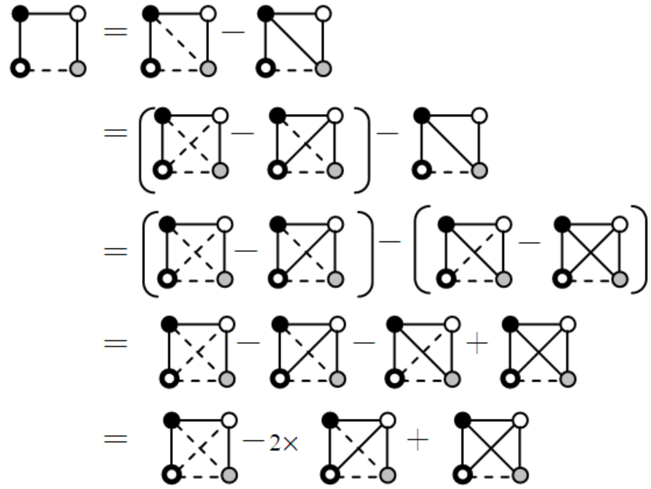


Fig. 12. Decomposing M into a linear combination of the selected basis.

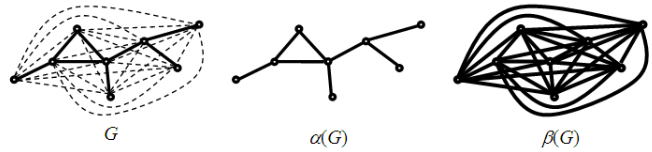


Fig. 13. A G (left) with its $\alpha(G)$ (middle) and $\beta(G)$ (right).

$$2 \times \left(\pi - \frac{3\sqrt{3}}{4}\right)\frac{\pi^2 r^6}{|A|^3} + \left(\pi - \frac{3\sqrt{3}}{4}\right)^2 \frac{\pi r^6}{|A|^3} = \frac{27}{16}\frac{\pi r^6}{|A|^3}.$$

D. Compute the probability of subgraphs with vertex size larger than four in RGGs

The subsection focuses on answering the question: In RGGs, what kinds of subgraphs whose probabilities can be exactly computed? This kind of graph is called *computable*. Here we define a new graph, called *Y-graphs*, which will be shown to be computable. First, we create two additional graphs $\alpha(G)$ and $\beta(G)$ from any class graph $G = (V, E_S, E_B)$ such that $\alpha(G) = (V, E_S)$ and $\beta(G) = (V, E_S \cup E_B)$. Fig. 13 shows an example.

A class graph $G = (V, E_S, E_B)$ is a *Y-graph* if it can be constructed according to the following three rules:

- R1: Any class graph with three vertices or any class graph in $\{G_1, G_2, \ldots, G_{11}\} - \{G_9, G_{11}\}$ is a Y-graph.
- R2: $G_i + G_j$ and $G_i - G_j$ are also Y-graphs if G_i and G_j are two disjoint Y-graphs.
- R3: G is a Y-graph if $\alpha(\phi(G))$ is a tree and $\beta(\phi(G))$ is a complete graph, where $\phi(G)$ is obtained from G by contracting a set of vertex-disjoint induced subgraphs, which are also Y-graphs.

Fig. 14 shows a Y-graph G with $\phi(G)$ by contracting a triangle. Given any Y-graph, Theorem 17 shows that its probability formula can be obtained exactly. To derive the result, we need the help of the following theorem.

Theorem 16 ([3]): A tree with two or more vertices has at least two leaves.

Theorem 17: Any Y-graph is computable.

Proof: Any class graph with three vertices or any class graph in $\{G_1, G_2, \ldots, G_{11}\} - \{G_9, G_{11}\}$ are computable as shown in Subsections IV-B and IV-C. If $\Pr(G_i)$ and $\Pr(G_j)$

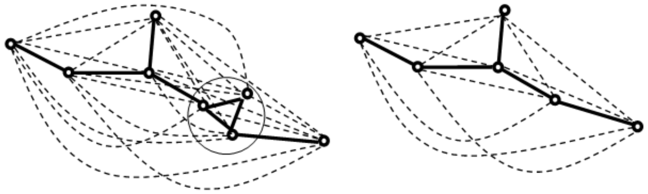


Fig. 14. A Y-graph G (left) and $\phi(G)$ (right), where $\alpha(\phi(G))$ is a tree and $\beta(\phi(G))$ is a complete graph.

are given, we have $\Pr(G_i + G_j) = \Pr(G_i) + \Pr(G_j)$ and $\Pr(G_i - G_j) = \Pr(G_i) - \Pr(G_j)$ if G_i and G_j are disjoint. The rest is the case for those Y-graphs constructed by applying R3.

Suppose that G is constructed by applying R3. Let S be the size of vertex set of $\phi(G)$. We will prove that G is computable by induction on S. When $S = 1$, then G is computable since it is a single vertex ($\Pr(G) = 1$) or a class graph with three vertices or a class graph in $\{G_1, G_2, \ldots, G_{11}\} - \{G_9, G_{11}\}$.

Since $\alpha(\phi(G))$ is a tree, $\alpha(\phi(G))$ must contain a leaf w by Theorem 16. The removal of w together with the edges incident with it from G results in G^*, which is computable according to the induction hypothesis.

If w is a vertex of G, we have only one solid line and $S-1$ broken lines incident to w due to the facts that $\alpha(\phi(G))$ is a tree and $\beta(\phi(G))$ is a complete graph; the existence of the unique solid line $e_j = (w, v)$ (where v is a vertex in G^*) only depends on whether the distance between w and v is less than r; therefore we have $\Pr(G) = \Pr(G^*) \times \Pr(E_j)$, indicating that G is computable.

Otherwise, w represents a Y-graph G_Y. Since every vertex in G^* connects to every vertex in G_Y with broken lines except one solid line $e_j = (x, y)$, where x (y) is a vertex in G^* (G_Y). Similarly, we have $\Pr(G) = Pr(G^*) \times \Pr(E_j) \times \Pr(G_Y)$; this also implies that G is also computable. ■

Finally, Corollary 18 is a trivial consequence of Theorem 17.

Corollary 18: Given a Y-graph $G = (V, E_S, E_B)$, if $\alpha(G)$ is a tree then $\Pr(G) = (\pi r^2/|A|)^{|V|-1}$.

V. Applications in Quantitative Analysis of Ad Hoc Networks

In the section, computing exact subgraph probability in RGGs is shown to be a practical tool for counting the number of induced subgraphs precisely, which explores fairly accurate quantitative property on topology of ad hoc networks.

A. Numbers of hidden-terminal pairs and dominating set

The number of hidden-terminal pairs has an impact on the performance of MAC layer in ad hoc networks. In [12], Khurana *et al.* have shown that if the number of hidden terminal pairs is small and when collisions are unlikely, the RTS/CTS exchange is a waste of bandwidth. On the other hand, if the number of hidden terminal pairs is large, RTS/CTS mechanism helps avoid collision. Moreover, the optimal value for the RTS_Threshold in IEEE 802.11 [12] depends on the number of hidden terminals.

Counting the number of hidden-terminal pairs in ad hoc networks becomes achievable with the help of the probability of hidden-terminal pair p_2 (Theorem 8). The expected number of hidden-terminal pairs is obtained in the next theorem.

Theorem 19: In a $\Psi(X_n, r, A)$, the number of expected number of hidden-terminal pairs equals $3 \times \binom{n}{3} \times \frac{3\sqrt{3}}{4}\pi r^4/|A|^2$.

Proof: Since each hidden-terminal pair consists of three distinct labeled vertices, we set S to be the selected three-vertex set. There are different combinations for selecting three from n vertices, and three different settings for labeling one from three as the center of the hidden-terminal pair (i.e. the internal node of the induced path with length 2). Therefore, the expected number of hidden-terminal pairs is $\binom{n}{3} \times 3 \times \Pr(G_S = p_2) = 3 \times \binom{n}{3} \times \left(\frac{3\sqrt{3}}{4}\right)\pi r^4/|A|^2$ by Table I. ■

Note that the number of the hidden terminal pairs grows as like $O(n^3 r^4)$ as the number of mobile nodes n or the communication range of each node r are increasing.

The probability of hidden-terminal pair p_2 has another application in estimating the number of cluster heads in ad hoc networks. The concept of dominating set in graph theory has been used for hierarchical routing and reducing broadcasting packets [4] in ad hoc networks. Finding the minimum size of dominating set confronts two huge obstacles. First, it is an NP-complete problem, which seems difficult to resolve efficiently. Second, it needs overheads to gather global knowledge of the network topology. Therefore, many approaches are proposed to find a dominating set with acceptable size. A simple distributed algorithm for selecting a dominating set in ad hoc networks has been proposed by Wu and Li [23]. Their algorithm selects the center of an induced path with length 2 (that is p_2) as a member of the desired dominating set. Similarly, with the help of the probability of hidden-terminal pair, the size of dominating set produced in their algorithm is estimated in the next theorem.

Theorem 20: Given a $\Psi(X_n, r, A)$, the expected size of dominating set $\gamma(G)$ is $n \times \left(1 - (1 - \left(\frac{3\sqrt{3}}{4}\right)\pi r^4/|A|^2)^{\binom{n-1}{2}}\right)$ if Wu and Li's dominating set algorithm is applied.

Proof: For each node (as a center), there are $\binom{n-1}{2}$ possible combinations for other two vertices to form a p_2. The node is not a member of the selected dominating set when every one pair of its neighbors with it does not form an induced path with length 2 (i.e., the p_2 subgraph). Therefore, the probability of a node being a member of the selected dominating set is $1 - (1 - \Pr(p_2))^{\binom{n-1}{2}} = \left(1 - (1 - \left(\frac{3\sqrt{3}}{4}\right)\pi r^4/|A|^2)^{\binom{n-1}{2}}\right)$ by Table I. For we have n possible centers, the expected size of dominating set is $n \times \left(1 - (1 - \left(\frac{3\sqrt{3}}{4}\right)\pi r^4/|A|^2)^{\binom{n-1}{2}}\right)$. ■

Evidently, the expected size of dominating set (as cluster heads or broadcasting nodes in ad hoc networks) is increasing as the number of deployed mobile nodes increases.

B. Number of exposed-terminal sets

In [21], Shukla *et al.* have mitigated the exposed terminal problem by identifying exposed terminal sets and scheduling concurrent transmissions whenever possible. By estimating the number of exposed-terminal sets in ad hoc networks, we can

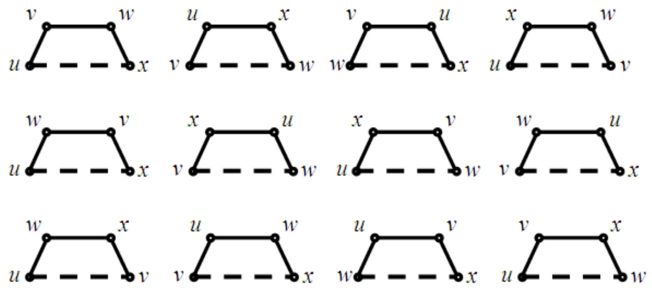


Fig. 15. Twelve different ways of labeling a subgraph M.

estimate the extent of system loss due to the exposed-terminal problem, and adopt similar techniques used in [21] to improve system performance.

To derive a tight bound of the number of exposed-terminal sets in a given RGG, we need the help of the subgraph probability of c_4 (an induced cycle of length four) whose upper bound has been derived in Theorem 13.

Theorem 21: In a $\Psi(X_n, r, A)$, the expected number of exposed-terminal sets is greater than $\left(\frac{2\times 3^4 - 9\sqrt{3}\pi}{8}\right)\binom{n}{4}\frac{\pi r^6}{|A|^3}$.

Proof: Counting the number of exposed-terminal sets is equivalent to counting the number of labeled subgraph M (Fig. 2). There are $\binom{n}{4}$ ways to select four from n nodes. The selected four distinct nodes has $\binom{4}{2} \times 2 = 12$ ways in forming a subgraph M (Fig. 15).

However, every class graph in the same row contains the same subgraph c_4 (an induced cycle of length four). To be accurate, we should avoid such duplicated counting for each selected four nodes. Therefore the number of exposed-terminal sets is equal to the number of labeled M graphs minus the number of the duplicated cycles (i.e., 3 (three duplicated counting in each row)×3 (three rows totally) for each selected four nodes). As a result, the expected number of exposed-terminal sets is

$$\binom{n}{4}\times\binom{4}{2}\times 2\times \Pr(G_s = M) - \binom{n}{4}\times 3\times 3\times \Pr(G_s = c_4)$$
$$= \binom{n}{4}\times\binom{4}{2}\times 2\times\frac{27}{16}\frac{\pi r^6}{|A|^3} - 9\times\binom{n}{4}\times \Pr(G_s = c_4)$$

(by the derived formula in Subsection IV-C-2-C)

$$= \frac{3^4}{4}\binom{n}{4}\frac{\pi r^6}{|A|^3} - 9\times\binom{n}{4}\times \Pr(G_s = c_4)$$
$$\geq \frac{3^4}{4}\binom{n}{4}\frac{\pi r^6}{|A|^3} - 9\times\binom{n}{4}\times\frac{\sqrt{3}}{8}\frac{\pi^2 r^6}{|A|^3} \quad \text{by Theorem 13}$$
$$= \left(\frac{2\times 3^4 - 9\sqrt{3}\pi}{8}\right)\binom{n}{4}\frac{\pi r^6}{|A|^3}.$$

■

Similarly, we conclude that the exposed-terminal sets grow as like $O(n^4 r^6)$ when the number of deployed mobile nodes n or the communication range r is increasing.

C. Numbers of triangle routes and backup routes

The triangle route problem is that messages transmit along a routing path from a through b to c without regard to the direct communication from a to c. Many existing routing protocols of ad hoc networks always select the shortest path for packet routing initially; however, when nodes arbitrarily move for a while, the selected path often suffers from the triangle route problem and thus is not the shortest one. The next theorem estimates the number of triangle routes in a given ad hoc network.

Theorem 22: In a $\Psi(X_n, r, A)$, the expected number of triangle routes equals $\binom{n}{3}\times\left(\pi - \frac{3\sqrt{3}}{4}\right)\pi r^4/|A|^2$.

Proof: The expected number of triangle routes can be obtained easily since we have computed $\Pr(c_3) = \left(\pi - \frac{3\sqrt{3}}{4}\right)\pi r^4/|A|^2$ (Table I); and there are $\binom{n}{3}$ ways to select three from n elements. Thus, we conclude that the expected number of triangle routes equals $\binom{n}{3}\times\Pr(c_3) = \binom{n}{3}\times\left(\pi - \frac{3\sqrt{3}}{4}\right)\pi r^4/|A|^2$. ■

Sometimes a triangle route can be treated in a different way. Once a route in ad hoc networks has been established, it can be maintained by a route maintenance procedure until either link failures occur or the transmission job is complete. Because routing information may not be available when a route request is received, the latency to determine a route can be quite significant. When the rate of topological changes in the network is sufficiently high, most existing protocols may not be able to react fast enough to maintain necessary routing. Some protocols try to replace a failed link (from a to b) with backup routes (from a through c to b) which is a triangle route [29], [32], [33]. Evidently, the number of triangle routes decides the suitability of these routing protocols as an effective recovery mechanism for link failures. Consequently, estimating the probability of link recovery from failures is essential to choose the time to activate the recovery mechanism. The next theorem estimates the probability of an arbitrary communication link (a, b) owning a two-hop backup route (from a through c to b).

Theorem 23: In a $\Psi(X_n, r, A)$, the probability of an arbitrary communication link owning a two-hop backup route is $(n-2)\times\left(\pi - \frac{3\sqrt{3}}{4}\right)r^2/|A|$.

Proof: Given two circles with the same radius r in a $\Psi(X_n, r, A)$, the expected overlapped area of two properly intersecting circles is $\left(\pi - \frac{3\sqrt{3}}{4}\right)r^2$ by Lemma 4. Subsequently the probability of a node located in the overlapped area of these two properly intersecting circles is $\left(\pi - \frac{3\sqrt{3}}{4}\right)r^2/|A|$. Finally, the probability of an arbitrary communication link owning a two-hop backup route is $(n-2)\times\left(\pi - \frac{3\sqrt{3}}{4}\right)r^2/|A|$. ■

Some protocols try to replace a failed link (from a to b) with a three-hop backup routes (from a through c and d to b) when a two-hop backup routes (from a through c to b) is not found.

The next theorem estimates the probability of an arbitrary communication link (a, b) owning a three-hop backup route.

Theorem 24: In a $\Psi(X_n, r, A)$, the probability of an arbitrary communication link owning a three-hop backup route is less than $(n-2)(n-3)\left(\frac{\sqrt{3}}{8}\right)\frac{\pi r^4}{|A|^2}$.

Proof: Suppose the interested communication link (a, b) exists for this moment. The three-hop backup route is from a through c and d to b. Note that this four nodes form an induced cycle with length four (i.e., c_4); otherwise, there would be a two-hop (shorter) backup route. There are at most $(n-2)(n-3)$ ways to select node c and d for a three-hop backup route. Given that $S = \{a, b, c, d\}$and $e_i = (a, b)$, the probability of an arbitrary communication link owning a three-hop backup route is $(n-2) \times (n-3) \times \Pr(G_S = c_4 \mid E_i) \leq (n-2) \times (n-3) \times \frac{\sqrt{3}}{8} \frac{\pi^2 r^6}{|A|^3} / \Pr(E_i) \leq (n-2) \times (n-3) \times \frac{\sqrt{3}}{8} \frac{\pi^2 r^6}{|A|^3} / ((\pi r^2)/|A|) = (n-2)(n-3) \left(\frac{\sqrt{3}}{8}\right) \frac{\pi r^4}{|A|^2}$ by Theorem 13. ■

VI. Conclusions

This work has proposed a novel paradigm for exactly computing numerous subgraph probabilities in RGGs in a systematical way. Applications in quantitative analysis on topology of ad hoc networks are also discussed. Future studies should devise an algorithm to recognize Y-graphs. Surely, we can apply the proposed paradigm for computing any k-vertex (for $k > 4$) subgraph probability exhaustively. However, it would be a tedious and challenging work. Consequently, quickly identifying a basis for complete class graph CG_k for a large k is also of priority concern.

Acknowledgment

The author is grateful to the anonymous referees for their helpful comments which have greatly improved the readability of this work.

References

[1] Christian Bettstetter, "On the minimum node degree and connectivity of a wireless multi-hop network," *MobiHoc*, 2002, pp. 80–91.
[2] B. Bollobas, *Random Graphs*, London: Academic Press, 1985.
[3] J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications*, Macmillan Press, 1976.
[4] Bevan Das and Vaduvur Bharghavan, "Routing in ad-hoc networks using minimum connected dominating sets," *IEEE International Conference on Communications*, 1997, pp. 376-380.
[5] J. Díaz, M. D. Penrose, J. Petit, and M. Serna, "Convergence theorems for some layout measures on random lattice and random geometric graphs," *Combinatorics, Probability, and Computing*, no. 6, pp. 489–511, 2000.
[6] J. Díaz, M. D. Penrose, J. Petit, and M. Serna, "Approximating layout problems on random geometric graphs," *J. Algorithms*, vol. 39, pp. 78–116, 2001.
[7] J. Díaz, J. Petit, and M. Serna, "Random geometric problems on $[0, 1]^2$," *Lecture Notes in Computer Science*, vol. 1518, Springer-Verlag, New York/Berlin, 1998.
[8] O. Dousse, P. Thiran, and M. Hasler, "Connectivity in ad-hoc and hybrid networks," *Infocom*, 2002.
[9] P. Erdös and A. Rénye, "On random graphs I," Publ. Math. Debrecen, vol. 6, pp. 290–297, 1959.
[10] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," *Stochastic Analysis, Control, Optimization and Applications*, pp. 547–566, 1998.
[11] Peter Hall, *Introduction to the Theory of Coverage Process*, John Wiley and Sons, New York, 1988.
[12] S. Khurana, A. Kahol, S. K. S. Gupta, and P. K. Srimani, "Performance evaluation of distributed co-ordination function for IEEE 802.11 wireless LAN protocol in presence of mobile and hidden terminals," *International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, 1999, pp. 40–47.
[13] S. Khurana, A. Kahol, and A. Jayasumana, "Effect of hidden terminals on the performance of the IEEE 802.11 MAC protocol," *Local Computer Networks Conference*, 1998.
[14] S.-J. Lee and M. Gerla, "AODV-BR: Backup routing in Ad hoc Networks," *IEEE Wireless Communications and Networking Conference*, 2000, vol. 3, pp. 1311–1316.
[15] Edgar M. Palmer, *Graphical Evolution: An Introduction to the Theory of Random Graphs*, New York: John Wiley and Sons, 1985.
[16] Mathew D. Penrose, *Random Geometric Graphs*, Oxford University Press, 2003.
[17] M. D. Penrose, "A strong law for the longest edge of the minimal spanning tree," *The Annals of Probability*, vol. 27, no. 1, pp. 246–260, 1999.
[18] M. D. Penrose, "The longest edge of the random minimal spanning tree," *The Annals of Applied Probability*, vol. 7, no. 2, pp. 340–361, 1997.
[19] M. D. Penrose, "On k-connectivity for a geometric random graph," *Random structures and Algorithms*, vol. 15, no. 2, pp. 145–164, 1999.
[20] Paolo Santi and Douglas M. Blough, "The critical transmitting range for connectivity in sparse wireless ad hoc networks," *IEEE Trans. Mobile Comput.*, vol. 2, no. 1, pp. 25–39, 2003.
[21] D. Shukla, L. Chandran-Wadia, and S. Iyer, "Mitigating the exposed node problem in IEEE 802.11 ad hoc networks," *International Conference on Computer Communications and Networks*, 2003, pp. 157–162.
[22] F. Tobagi and L. Kleinrock, "Packet switching in radio channels, Part II-The hidden terminal problem in carrier sense multiple access and the busy tone solution," *IEEE Trans. Commun.*, vol. COM-23, no. 12, pp. 1417–1433, 1975.
[23] J. Wu and H. Li, "Domination and its application in ad hoc wireless networks with unidirectional links," *International Conference on Parallel Processing*, 2000, pp. 189–197.
[24] F. Xue and P. R. Kumar, "The number of neighbors needed for connectivity of wireless networks," *Wireless Networks*, vol. 10, pp. 169–181, 2004.
[25] K. J. Yang, *On the Number of Subgraphs of a Random Graph in* $[0, 1]^d$, Unpublished D. Phil. thesis, Department of Statistics and Actuarial Science, University of Iowa, 1995.
[26] L.-H. Yen and Chang Wu Yu, "Link probability, network coverage, and related properties of wireless ad hoc networks," *The 1st IEEE International Conference on Mobile Ad-hoc and Sensor Systems*, 2004, pp. 525–527.
[27] L.-H. Yen, Chang Wu Yu, and Yang-Min Cheng, "Expected k-coverage in wireless sensor networks," *Ad Hoc Networks*, vol. 5, no. 4, pp. 636–650, 2006.
[28] Chang Wu Yu, "Quantitative analysis of multi-hop wireless networks using a novel paradigm," *International Symposium on Combinatorics, Algorithms, Probabilistic and Experimental Methodologies* (ESCAPE 2007).
[29] Chang Wu Yu, Tung-Kuang Wu, and Rei Heng Cheng, "A low overhead dynamic route repairing mechanism for mobile ad hoc networks," *Computer Communications*, vol. 30, pp. 1152–1163, 2007.
[30] Chang Wu Yu and Li-Hsing Yen, "Computing subgraph probability of random geometric graphs: Quantitative analyses of wireless ad hoc networks," *25th IFIP WG 6.1 International Conference on Formal Techniques for Networked and Distributed Systems*, 2005.
[31] Chang Wu Yu, L.-H. Yen, and Yang-Min Cheng, "Computing subgraph probability of random geometric graphs with applications in wireless ad hoc networks," *Tech. Rep., CHU-CSIE-TR-2004-005, Chung Hua University, R.O.C.*
[32] Chang Wu Yu, Li-Hsing Yen, Kun-Ming Yu, and Zhi Pin Lee, "An ad hoc routing protocol providing short backup routes," *The Eighth IEEE International Conference on Communication Systems*, 2002, Singapore, pp. 1052–1056.
[33] Kun-Ming V. Yu, Shi-Feng Yand, and Chang Wu Yu, "An ad hoc routing protocol with multiple backup routes," *Proc. IASTED International Conference Networks, Parallel and Distributed Processing, and Applications*, 2002, pp. 75–80.
[34] http://mathworld.wolfram.com/.

Chang Wu Yu was born in Taoyuan, Taiwan, R. O. C. in 1964. He received the BS degree from Soochow University in 1985, MS degree from National Tsing Hua University in 1989, and Ph.D. degree from National Taiwan University in 1993, all in computer sciences. From 1995 to 1998, he was an Associate Professor at the Department of Information Management, Ming Hsin Institute of Technology. In 1999, he joined the Department of Computer Science & Information Engineering, Chung Hua University. His current research interests include graph algorithms and wireless networks. Dr. Yu received best paper awards at 2008 ACM International Conference on Sensor, Ad Hoc, and Mesh Networks and at both 2004 and 2007 Mobile Computing Workshop.