The number of possibilities for random dating

Aaron Abrams* and Rod Canfield† and Andrew Granville‡

Abstract

Let G be a regular graph and H a subgraph on the same vertex set. We give surprisingly compact formulas for the number of copies of H one expects to find in a random subgraph of G.

1 Introduction

There are n boys and n girls who apply to a computer dating service, which randomly picks a boy and a girl and then introduces them. It does this again and again until everyone of the boys and girls has been introduced to at least one other person. The service then organizes a special evening at which everyone dates someone to whom they have been previously introduced. In how many different ways can all of the boys and girls be matched up?

Translating this question into the language of graph theory, we select edges at random from the complete bipartite graph $K_{n,n}$ until the subgraph created by these edges, G, has minimum degree 1. We then ask how many perfect matchings (subgraphs on the same vertex set where every vertex has degree 1) are contained in G?

We know that, with probability $1 - o(1)$, there will be at least one possible matching [1] at the time that the last person is finally introduced to someone, and that there will have been a total of $\sim n \log n$ introductions made [2]. Now, note that if the jth girl has been introduced to b_j boys then the number of possible matchings is certainly $\leq b_1 b_2 \ldots b_n \leq (k/n)^n$ where $k = b_1 + b_2 + \ldots + b_n$, by the arithmetic-geometric mean inequality. In particular we expect no more than $(\{1 + o(1)\} \log n)^n$ ways of matching up the boys and girls, by the time the last person is finally introduced to someone. It thus came as a surprise to us, when we did the calculation, to discover that one expects there to be far more, in fact something like $(n/4\mathrm{e})^n$ ways of matching up the boys and girls!

The fault in the above reasoning comes in believing that what we expect to happen in the most likely case (in which there have been $\sim n \log n$ introductions

*Department of Mathematics and Computer Science, Emory University, Atlanta GA 30322, USA, abrams@mathcs.emory.edu

†Department of Computer Science, University of Georgia, Athens GA 30602, USA, erc@cs.uga.edu

‡Département de Mathématiques et Statistique, Université de Montréal, CP 6128 succ Centre-Ville, Montréal QC H3C 3J7, Canada, andrew@dms.umontreal.ca

made by the time the last person is finally introduced to someone) yields what we expect to happen in general. In fact the largest contribution to the expectation takes place in the very rare situation that the computer dating service somehow neglects to make any introductions involving one sad participant until about $n^2/2$ introductions have already been made involving only the others.

To see this, first note that the probability that the nth boy has not been introduced to anyone by the time $k = [n^2/2]$ random introductions have been made is[1]

$$\frac{n^2-n}{n^2} \cdot \frac{n^2-n-1}{n^2-1} \cdot \frac{n^2-n-2}{n^2-2} \cdots \frac{n^2-n-k+1}{n^2-k+1} \sim \frac{\mathrm{e}^{-1/2}}{2^n}.$$

With probability $1 + o_n(1)$, each person other than the nth boy has, by now, been introduced to $\sim n/2$ people (here $o_n(1)$ represents a function that $\to 0$ as $n \to \infty$). Suppose that the $(k+1)$st introduction involves the nth boy and girl. For any of the $(n-1)!$ possible matchings involving all of the boys and girls in which the nth boy and girl are matched up, the probability that each of the pairings in that matching has already been introduced is close to $1/2$, so we might guess that the probability that the matching can occur with the introductions already made is around $1/2^{n-1}$. Thus we might expect that the contribution to the expectation when $k = [n^2/2]$ is around $(n-1)! \times (\mathrm{e}^{-1/2}/2^n) \times (1/2^{n-1}) \approx (n/4\mathrm{e})^n$.

The main point of this article is to make this discussion precise with complete proofs, and in some generality. The exact formulas obtained are surprisingly compact. Let us mention that in [3] Janson has proven that for a random graph on n vertices with m edges various subgraph counts (spanning trees, Hamilton cycles, and perfect matchings) are asymptotically normal for a suitable range of m.

2 Statement of results

Let G_ω be a random subgraph of $K_{n,n}$ formed by randomly adding edges until every vertex has degree at least one. Various things are known about such G_ω; for instance G_ω is almost surely connected and G_ω almost surely contains a matching [1]. We ask how many matchings does G_ω contain?

One can ask the same question for random subgraphs G_ω of K_{2n}, or indeed of any other graph G. Likewise instead of matchings one could count the expected number of occurrences of any other prescribed subgraph H on the same vertex set as G.

Remark. Here and henceforth, whenever we refer to counting the occurrences of a prescribed subgraph H, we always have in mind a *spanning* subgraph. In particular, the minimum degree of H is 1.

Of course, before any copy of H can appear as a subgraph of G_ω, it is necessary that the minimum degree of G_ω be at least as great as the minimum degree of

[1] For this and later estimates, note that $\log(A(A-1)\dots(A-(n-1))/A^n) = -n^2/2A + O(1/n)$ for $A \gg n^2$.

H. Note also that G itself may not contain any subgraphs isomorphic to H; thus it makes sense to investigate the proportion of subgraphs of G which are isomorphic to H that actually occur as a subgraph of G_ω. Define

$$\mathbb{E}(H \subseteq G) := \mathbb{E}\left(\frac{\#\{J \subseteq G_\omega \ : \ J \cong H\}}{\#\{J \subseteq G : J \cong H\}}\right).$$

Thus let G and H be graphs with $|V(G)| = |V(H)|$, and let $\delta \geq 1$ be the minimum degree of H. Let Ω be the set of permutations of the m edges of G. For $\omega = (e_{\omega(1)}, \ldots, e_{\omega(m)}) \in \Omega$, let $G_\omega^{(i)}$ be the graph with vertex set $V(G)$ and edge set $\{e_{\omega(1)}, \ldots, e_{\omega(i)}\}$, and let

$$k(\omega) = \min\{i : G_\omega^{(i)} \text{ has minimum degree } \delta\}.$$

Then define $G_\omega = G_\omega^{(k(\omega))}$. The question now becomes: for random $\omega \in \Omega$, what fraction of those subgraphs of G which are isomorphic to H are contained in G_ω?

Our main theorem answers this question in the case that G is regular.

Theorem 1 *Suppose G is d-regular, and let H be a subgraph of G with the same vertex set, of minimal degree $\delta \geq 1$. Let h be the number of edges of H, and let $\Delta = d - \delta$. Then*

$$\mathbb{E}(H \subseteq G) = \frac{2}{\binom{h+\Delta}{h}} - \frac{1}{\binom{h+2\Delta}{h}}.$$

We can apply this to our original problem.

Corollary 1 *In a random subgraph of $K_{n,n}$ formed by adding edges until each vertex has degree at least one, the expected number of complete matchings is*

$$n!\left(\frac{2}{\binom{2n-1}{n}} - \frac{1}{\binom{3n-2}{n}}\right).$$

Proof. In the notation of the theorem, $d = n$, $\delta = 1$, $h = n$, and $\Delta = n - 1$. Apply the theorem, and note that there are $n!$ distinct complete matchings in $K_{n,n}$. □

If boys can date boys, and girls can date girls, we can still ask the same questions, and again obtain what is to us a surprising answer.

Corollary 2 *Let G be any d-regular graph on $2n$ vertices, and form G_ω by randomly choosing edges of G until each vertex has degree at least one. Of all the complete matchings in G, the fraction expected to occur in G_ω is*

$$\frac{2}{\binom{n+d-1}{n}} - \frac{1}{\binom{n+2d-2}{n}}.$$

Corollary 3 *Let G be a d-regular graph on n vertices, and form G_ω by randomly choosing edges of G until each vertex has degree at least two. Of all the Hamiltonian cycles in G, the fraction expected to occur in G_ω is*

$$\frac{2}{\binom{n+d-2}{n}} - \frac{1}{\binom{n+2d-4}{n}}.$$

The proof of the main theorem uses a simple inclusion-exclusion argument. It is not hard to prove a result for more general graphs G and H as indicated in the proof, though this would be complicated to state. In [4] McDiarmid proves these results in the special case that $G = K_n$ though with more examples of desired subgraphs (that is, other than just for matchings and for Hamiltonian cycles). In our proof, in section 3, we are able to take any G, and any desired subgraph H, so long as it has as many vertices as G. Our proof has similarities to that in [4], though it is not entirely clear how to generalize [4] directly to recover the general results presented in section 3.

There is a hypergraph version whose proof is essentially the same but whose statement is correspondingly more complicated. Thus we will state two special cases.

First, let $G = K_{n,n,\ldots,n}$ be a complete r-partite graph with vertex set $V(G)$. Let E be the set of r-element subsets of $V(G)$ which span K_r subgraphs of G (so $|E| = n^r$). Form an r-uniform hypergraph (on the same vertex set as G) by choosing random elements of E as edges until each vertex is contained in at least one edge; i.e. choose a permutation ω of E and let G_ω be defined as it was before.

A *matching* in such a hypergraph G_ω is a set M of edges such that each vertex of G is in exactly one edge of M. (Equivalently a matching is a copy, inside G_ω, of the hypergraph H consisting of n disjoint edges, each of size r.)

Theorem 2 *With $G = K_{n,n,\ldots,n}$ the complete r-partite graph, and with G_ω as above, the expected number of matchings in G_ω is*

$$(n!)^{r-1} \sum_{i=1}^{r} (-1)^{i-1} \frac{\binom{r}{i}}{\binom{n^r-(n-1)^i n^{r-i}+n-1}{n}}.$$

The factor $(n!)^{r-1}$ is the total number of matchings in G. The case $r = 2$ agrees with Corollary 1. For $r \geq 3$ we have

$$n^r - (n-1)^i n^{r-i} + n - 1 = i n^{r-1}(1 + O(1/n)),$$

so that

$$\binom{n^r - (n-1)^i n^{r-i} + n - 1}{n} \asymp (i n^{r-1})^n / n!$$

Thus the $i = 1$ term is the main term in the sum, and so the sum can be estimated as

$$\{r c_r + O_r(1/n)\}(n!)^r / n^{(r-1)n},$$

where $c_3 = \mathrm{e}^{-1/2}$ and $c_r = 1$ for $r \geq 4$. A more crude estimate is $\asymp_r n^{O_r(1)}(n/\mathrm{e}^r)^n$ (where "$O_r(1)$" is in place of a function that is bounded in terms of r only, and where "$A \asymp_r B$" means that A/B is bounded above and below by positive constants that depend only on r).

For our second hypergraph theorem, let G be the complete graph K_{rn}. Let E be the set of all r-element subsets of $V(G)$. We form an r-uniform hypergraph G_ω by taking edges from E in some order ω, stopping when each vertex has degree at least one. Again, we find the expected number of matchings in such a hypergraph G_ω.

Theorem 3 *With $G = K_{rn}$ and with G_ω as above, the expected number of matchings in G_ω is*

$$\frac{(rn)!}{(r!)^n n!}\sum_{i=1}^{r}(-1)^{i-1}\frac{\binom{r}{i}}{\binom{\binom{nr}{r}-\binom{nr-i}{r}+n-1}{n}}.$$

Again, the factor preceding the sum is the total number of matchings in G. The case $r = 2$ agrees with Corollary 2 with $G = K_{2n}$. For $r \geq 3$ we have that if $i \geq 2$ then $\binom{nr}{r} - \binom{nr-i}{r} \geq (\binom{nr}{r} - \binom{nr-1}{r})(2 - \frac{(r-1)}{(nr-1)})$, and thus the $i \geq 2$ terms have magnitude $\ll 1/2^n$ times the magnitude of the $i = 1$ term. Therefore the sum can be estimated as $\{rc'_r + O_r(1/n)\}(((n-1)r)!n)^n/(rn)!^{n-1}$ where $c'_3 = \mathrm{e}^{-1/9}$ and $c'_r = 1$ for $r \geq 4$. A more crude estimate is $\asymp_r n^{O_r(1)}(n/\mathrm{e}^r)^n$.

The proofs of theorems 2 and 3 are the obvious generalizations of the proof of theorem 1. Thus we have restricted ourselves to sketching the proof of theorem 2, and leaving the proof of theorem 3 to the enthusiastic reader.

3 Proof of theorem 1

We are given graphs G and H with $V(G) = V(H)$ and where $\min_{v\in V(G)} \deg_G(v) \geq \min_{v\in V(G)} \deg_H(v)$. For now we need not assume that G is regular. We use the notation δ, Ω, $k(\omega)$, G_ω, and h as defined in the previous section.

By definition

$$\mathbb{E}(H \subseteq G) = \frac{1}{\#\{J \subseteq G : J \cong H\}} \sum_{\substack{J\subseteq G\\ J\cong H}} \Pr(e_1, \ldots, e_{k(\omega)} \supseteq J :\ \omega \in \Omega).$$

Now we fix $J \subseteq G$ with $J \cong H$ and evaluate the probability that $J \subseteq e_1, \ldots, e_{k(\omega)}$ as we vary over $\omega \in \Omega$. Fix ω, and let u and v be the endpoints of the edge $e_{k(\omega)}$. By definition of $k(\omega)$, either $\deg_{G_\omega}(u) = \delta$ or $\deg_{G_\omega}(v) = \delta$ (or both).

Now, $J \subseteq G_\omega$ with $e_{k(\omega)} = uv$ and $\deg_{G_\omega}(u) = \delta$ if and only if certain edges of G appear in the correct order in ω:

- first, the edges of J other than uv (in some order);
- then uv;

- then the edges of G which contain u other than those in J.

Note that the location in ω of the other edges of G is irrelevant.

The probability that the edges appear in this order is

$$\frac{(h-1)!1!(\deg_G(u)-\delta)!}{(h+\deg_G(u)-\delta)!},$$

where h is the number of edges of H.

A similar count holds if $\deg_{G_\omega}(v)=\delta$. However, we have now double counted the cases in which $\deg_{G_\omega}(u)=\delta=\deg_{G_\omega}(v)$. By similar reasoning, the probability of this is

$$\frac{(h-1)!1!(\deg_G(u)+\deg_G(v)-2\delta)!}{(h+\deg_G(u)+\deg_G(v)-2\delta)!}.$$

Thus, given that the last edge $e_{k(\omega)}$ is the edge uv, we have

$$\Pr(e_1,\ldots,e_{k(\omega)}\supseteq J,\ \text{ and } e_{k(\omega)}=uv:\ \omega\in\Omega)=\frac{1}{h}\left(\frac{2}{\binom{h+\Delta(u)}{h}}-\frac{1}{\binom{h+\Delta(u)+\Delta(v)}{h}}\right),$$

where $\Delta(u)=\deg_G(u)-\delta$. Note that the only dependence of the right side on J is the choice of the edge uv from J. We deduce immediately that

$$\Pr_{\omega\in\Omega}(e_1,\ldots,e_{k(\omega)}\supseteq J:\ \omega\in\Omega)=\sum_{uv\in J}\frac{1}{h}\left(\frac{2}{\binom{h+\Delta(u)}{h}}-\frac{1}{\binom{h+\Delta(u)+\Delta(v)}{h}}\right).$$

Now if G is d-regular then $\Delta(u)=d-\delta$ for all vertices u, and our sum is over h edges, so that

$$\Pr_{\omega\in\Omega}(e_1,\ldots,e_{k(\omega)}\supseteq J)=\frac{2}{\binom{h+\Delta}{h}}-\frac{1}{\binom{h+2\Delta}{h}}$$

unconditionally. $\square$

Remark: The proof above yields, when G is not regular, that the expected number of copies of H in G_ω is

$$\frac{1}{h}\sum_{uv\in E(G)}\left(\frac{2}{\binom{h+\Delta(u)}{h}}-\frac{1}{\binom{h+\Delta(u)+\Delta(v)}{h}}\right)\#\{J\subseteq G:\ J\cong H,\ \text{and } uv\in J\}.$$

4 Proof of theorem 2

Denote the vertices of $K_{n,\ldots,n}$ by $v_{i,j}$ where $1\le i\le r$ and $1\le j\le n$. Each edge of E has one vertex $v_{i,j(i)}$ for each i; we abbreviate the edge as $(j(1),\ldots,j(r))$.

To prove theorem 2 we apply the same technique as in the proof of theorem 1. The expected number of matchings in $G_\omega=\{e_1,\ldots,e_{k(\omega)}\}$ is the total number of possible matchings (which is $(n!)^{r-1}$) times the probability that any given matching occurs in G_ω. We may compute the latter probability for the "diagonal" matching $M = \{(1,\dots,1),\dots,(n,\dots,n)\}$ as follows:

$$\begin{aligned}\Pr(G_\omega \supseteq M) &= n\Pr\left(G_\omega \supseteq M \;\text{ and } e_{k(\omega)} = (n,\dots,n)\right)\\ &= n\sum_{i=1}^{r}(-1)^{i-1}\binom{r}{i}\Pr(G_\omega \supseteq M,\; e_{k(\omega)} = (n,\dots,n),\\ &\qquad\qquad \text{and } v_{1,n},\dots,v_{i,n} \;\text{ all have degree one in } G_\omega),\end{aligned}$$

where the last equality follows by inclusion-exclusion, and the symmetry of G. We can determine the latter probability as before, namely, the event occurs if and only if the relevant edges occur in ω in the correct order:

- first, the edges $\{(i,\dots,i) \;:\; 1 \le i \le n-1\}$ (in some order);
- then $(n,\dots,n)$;
- then any edge (other than $(n,\dots,n)$) containing $v_{j,n}$ for some j, $1 \le j \le i$.

Again, the other edges can occur anywhere in ω. Thus the desired probability is

$$\frac{(n-1)!(n^r-(n-1)^i n^{r-i}-1)!}{(n^r-(n-1)^i n^{r-i}+n-1)!} = \frac{1}{n}\binom{n^r-(n-1)^i n^{r-i}+n-1}{n}^{-1}.$$

This establishes theorem 2.

5 Contributions to the expectation in Theorem 1

To better understand the result in Theorem 1, we now determine the contribution to the expectation of the sum, over all ω with $k(\omega) = k$, of the number of copies of H contained in $e_1,\dots,e_k$. Let E be the total number of edges in G. Just as in the proof of Theorem 1 the main term is given by $2h$ times the number of sets $e_1,\dots,e_k$ which contain J, with $e_k = uv$, where the other edges of G which contain u lie outside $e_1,\dots,e_k$. Once these edges are chosen, the number of possibilities for the $k-h$ edges in $\{e_1,\dots,e_k\}\setminus J$ is $\binom{E-h-\Delta}{k-h}$. The number of possible orderings of the edges in G given the above restrictions is $(k-1)!1!(E-k)!$. By an analogous calculation when u and v both have degree δ in G_ω, we have that the sum, over all ω with $k(\omega) = k$, of the number of copies of H contained in $e_1,\dots,e_k$, is

$$h(k-1)!1!(E-k)!\left\{2\binom{E-h-\Delta}{k-h} - \binom{E-h-2\Delta}{k-h}\right\}.$$

If $h = o(k)$ and $k = o(E)$ then this is $E!(k/E)^h \exp(O(k^2/E + h^2/k))$. It is maximized when k is a little larger, namely when $k \sim hE/(\Delta + h)$ (and note that there are no more than $E - h + 1$ possible values for k). Since $\Delta \leq d - 1 \leq |V(G)| - 2 = |V(H)| - 2 < \delta|V(H)| \leq 2h$, we have $E \geq hE/(\Delta + h) \geq E/3$.

In the case considered in Corollary 1, $h = n, \Delta = n - 1$ and $E = n^2$, so the main contribution to the expectation occurs when $k \sim n^2/2$, far more edges than when $k \sim n \log n$ which is the value we expect for $k(\omega)$. Likewise, for matchings in K_{2n} the biggest contribution occurs with $k = 2n^2/3$ rather than the expected $k \sim n \log n$, and for Hamiltonian cycles in K_n the biggest contribution comes from $k = n^2/4$, not $k \sim (n/2) \log n$.

References

[1] Béla Bollobás, *Random Graphs*, Cambridge Studies in Advanced Mathematics **73** (2001).

[2] Paul Erdős and Alfréd Rényi, *On random graphs I*, Publ. Math. Debrecen **6** (1959) pp. 290–297.

[3] Svante Janson, *The numbers of spanning trees, Hamilton cycles, and perfect matchings in a random graph*, Combin. Probab. Comput. **3** (1994) pp. 97–126.

[4] Colin McDiarmid, *Expected Numbers at Hitting Times*, J. Graph Theory **15** (1991) pp. 637–648.