Oberwolfach Report: Combinatorics and Probability, April 2013

Large deviations in random graphs

Yufei Zhao
(joint work with Eyal Lubetzky)

The goal of this talk is to answer the following question of Chatterjee and Varadhan [3].

Question 1. Fix $0 < p < r < 1$. Let G_n be an instance of the Erdős-Rényi random graph $G(n,p)$ conditioned on the rare event of having at least $\binom{n}{3}r^3$ triangles. For large n, is G_n close in cut distance to a typical $G(n,r)$?

By cut distance here we mean the quantity

$$\delta_\square(G_n, r) = \max_{A,B\subseteq V(G_n)} \frac{1}{n^2}|e(A,B) - r|A||B||.$$

The problem of large deviations in random graphs has a long history. A representative example that drew much interest is the upper tail problem of estimating the probability that $G(n,p)$ has at least $(1+\eta)\binom{n}{3}p^3$ triangles, where η is fixed and p is allowed to vary with n. Janson, Oleszkiewicz, and Ruciński [5] and Kim and Vu [6] developed powerful techniques for proving concentration bounds for this problem. In recent breakthroughs independently by Chatterjee [1] and DeMarco and Kahn [4] they showed that the upper tail probability for triangle counts is $e^{-\Theta_\eta(n^2p^2\log(1/p))}$ when $p \geq \log n/n$, thereby determining the correct order in the exponent.

We consider the case of constant p, with the goal of determining the constant in the exponent of the upper tail probability. More precisely we are interested in the quantity

$$\text{Rate} := -\lim_{n\to\infty} \frac{1}{\binom{n}{2}} \log \mathbb{P}(G(n,p) \text{ has at least } \tbinom{n}{3}r^3 \text{ triangles}).$$

Observe that if the number of edges in $G(n,p)$ deviates to $\binom{n}{2}r$, with the edges uniformly distributed, then one has the desired triangle count deviation. This gives an upper bound on the rate

$$\begin{aligned}\text{Rate} \leq h_p(r) &:= r\log\frac{r}{p} + (1-r)\log\frac{1-r}{1-p}\\ &= -\lim_{N\to\infty}\frac{1}{N}\log\mathbb{P}(\text{Binom}(N,p) \geq Nr).\end{aligned}$$

However, there could be other "reasons" for generating too many triangles. One has the following dichotomy, with names borrowed from statistical physics, for the question: *what is the most likely reason for having too many triangles?*

Replica symmetric phase: Too many edges, uniformly distributed.
In this case Rate $= h_p(r)$ and the answer to Question 1 is YES.
Symmetry breaking phase: Some other configuration (e.g., a large clique)
In this case Rate $< h_p(r)$ and the answer to Question 1 is NO.

The answer turns out to depend on (p,r). Chatterjee and Dey [2] showed, using Stein's method, that the region to the right of the dashed curve in Figure 1 belongs to the replica symmetric phase. Chatterjee and Varadhan [3] developed a new framework using graph limits

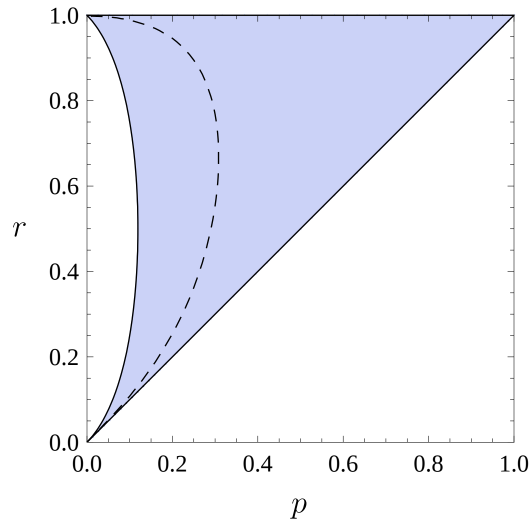

FIGURE 1. Phase diagram for the upper tail of triangle counts. Shaded region is the replica symmetric phase; the region to its left is the symmetry breaking phase. Previous results [2, 3] established replica symmetry to the right of the dashed curve.

and rediscovered the same replica symmetric region as [2], and furthermore they showed that for each r, one must enter the symmetry breaking phase for sufficiently small p.

Applying the framework of Chatterjee and Varadhan, which reduced the determination of the large deviation rate to an extremal problem for graphons, we identified the complete replica symmetric phase, plotted as the shaded region in Figure 1.

Theorem 2. [7] *The replica symmetric phase for upper tail large deviations in triangle counts is given by* $\{(p, r) : (1 + (r^{-1} - 1)^{1/(1-2r)})^{-1} \leq p < r\}$.

Furthermore, we identified the replica symmetric phase when the triangle is replaced by any d-regular graph. It turns out that the phase diagram depends only on d. The boundary curves are shown in Figure 2. We also considered the upper tail problem of having largest eigenvalue of $G(n, p)$ being at least nr (it's typically concentrated near np). The resulting phase diagram coincides with triangle count large deviations.

In the talk I explained the Chatterjee-Varadhan framework, our solution to the problem, as well as many open problems that remain.

REFERENCES

[1] S. Chatterjee. The missing log in large deviations for triangle counts. *Random Structures Algorithms*, 40(4):437–451, 2012.

[2] S. Chatterjee and P. S. Dey. Applications of Stein's method for concentration inequalities. *Ann. Probab.*, 38(6):2443–2485, 2010.

[3] S. Chatterjee and S. R. S. Varadhan. The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.*, 32(7):1000–1017, 2011.

[4] B. DeMarco and J. Kahn. Upper tails for triangles. *Random Structures Algorithms*, 40(4):452–459, 2012.

[5] S. Janson, K. Oleszkiewicz, and A. Ruciński. Upper tails for subgraph counts in random graphs. *Israel J. Math.*, 142:61–92, 2004.

[6] J. H. Kim and V. H. Vu. Divide and conquer martingales and the number of triangles in a random graph. *Random Structures Algorithms*, 24(2):166–174, 2004.

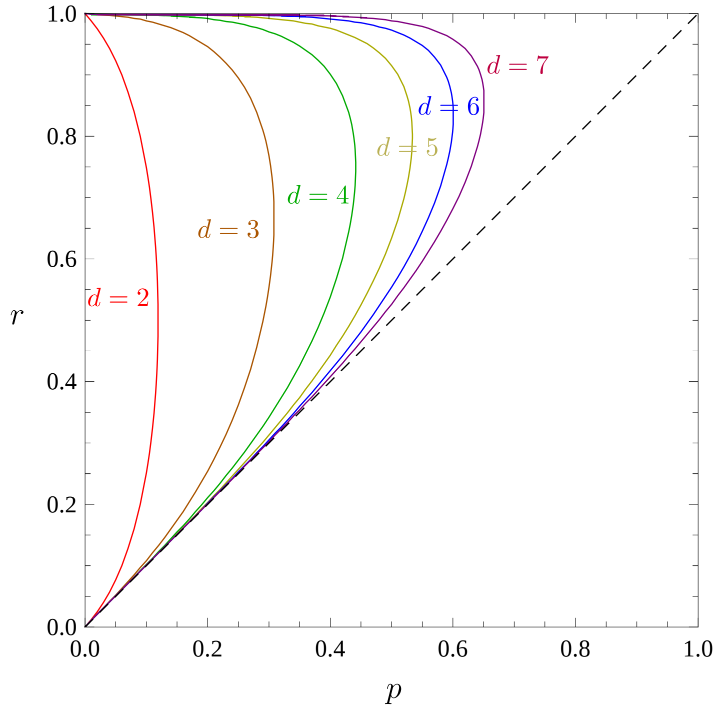

FIGURE 2. The phase boundary for counts of d-regular fixed subgraphs in $G(n, p)$.

[7] E. Lubetzky and Y. Zhao. On replica symmetry of large deviations in random graphs. Preprint arXiv:1210.7013.