ELSEVIER

A random triadic process

Dániel Korándi [1]

Department of Mathematics
ETH Zurich
Zurich, Switzerland

Yuval Peled [2]

School of Computer Science and Engineering
The Hebrew University of Jerusalem
Jerusalem, Israel

Benny Sudakov [3]

Department of Mathematics
ETH Zurich
Zurich, Switzerland

Abstract

Given a random 3-uniform hypergraph $H = H(n, p)$ on n vertices where each triple independently appears with probability p, consider the following *graph* process. We start with the star G_0 on the same vertex set, containing all the edges incident to some vertex v_0, and repeatedly add an edge xy if there is a vertex z such that xz and yz are already in the graph and $xyz \in H$. We say that the process *propagates* if all the edges are added to the graph eventually. In this paper we prove that the threshold probability for propagation is $p = \frac{1}{2\sqrt{n}}$. We also show that $p = \frac{1}{2\sqrt{n}}$ is an

upper bound for the threshold probability that a random 2-dimensional simplicial complex is simply-connected.

1 Introduction

The principle of *triadic closure* is an important concept in social network theory (see e.g. [5]). Roughly speaking, it says that when new friendships are formed in a social network, it is more likely to occur between two people sharing a common friend, thus "closing" a triangle, than elsewhere. We will consider a simplistic model of the evolution of a social network, where friendships can only be formed through a common friend, and triadic closure eventually occurs at any triangle with probability p, independently of other triangles. We refer to this process as the *triadic process*.

Formally, let $H = H(n, p)$ be a random 3-uniform hypergraph on $[n]$ where each triple independently appears with probability p. The triadic process is the following *graph* process. We start with the star G_0 on the same vertex set $[n]$, containing all the edges incident to some vertex v_0, and repeatedly add any edge xy if there is a vertex z such that xz and yz are already in the graph and $xyz \in H$. We say that the process *propagates* if all the edges are added to the graph eventually. It is easy to see that this event does not depend on the order the edges are added in. In this paper we prove that the threshold probability for propagation is $\frac{1}{2\sqrt{n}}$.

Theorem 1.1 *Suppose $p = \frac{c}{\sqrt{n}}$, for some constant $c > 0$. Then,*

(i) *If $c > \frac{1}{2}$, then the triadic process propagates whp.*

(ii) *If $c < \frac{1}{2}$, then the triadic process stops at $O(n\sqrt{n})$ edges whp.*

As usual, we say that some property holds *with high probability* or *whp* if it holds with probability tending to 1 as n tends to infinity.

Randomized graph processes have been intensively studied in the past decades. One notable example is the triangle-free process, originally motivated

by the study of the Ramsey number $R(3, n)$ (see e.g. [6]). In this process the edges are added one by one at random as long as they do not create a triangle in the graph. The triadic process is a slight variant of this, with a very similar nature. Indeed, our analysis makes good use of the tools developed by Bohman [2] when he applied the differential equation method to track the triangle-free process. Several other related processes were also analyzed using differential equations, e.g. [3]. For more information about this method we refer the interested reader to the excellent survey of Wormald [8].

Coja-Oghlan, Onsjö and Watanabe [4] investigated a similar kind of closure while analyzing connectivity properties of random hypergraphs. They say that a 3-hypergraph is propagation connected if its vertices can be ordered in some way $v_1, \ldots, v_n$ so that each v_i $(i \geq 3)$ forms a hyperedge with two preceding vertices. They obtain the threshold probability for the propagation connectivity of $H(n, p)$ up to a small multiplicative constant. Using this directed notion of connectivity, our problem asks when the random 3-hypergraph on the *line graph* of K_n is propagation connected from the star.

Our main motivation for considering the triadic process comes from the theory of random 2-dimensional simplicial complexes. A simplicial 2-complex on the vertex set V is a set family $Y \subseteq \binom{V}{\leq 3}$ closed under taking subsets. The dimension of a simplex $\sigma \in Y$ is defined to be $|\sigma| - 1$. We use the terms vertices, edges and faces for 0, 1 and 2-dimensional simplices, respectively. The 1-skeleton of a 2-complex is the subcomplex containing its vertices and edges.

The Linial–Meshulam model of random simplicial complexes, introduced in [7], is a generalization of the Erdős–Rényi random graph model and has been studied extensively in recent years. The random 2-complex $Y_2(n, p)$ is defined to have the complete 1-skeleton, i.e. all vertices and edges, and each of the faces independently with probability p. The study of random complexes involves both topological invariants and combinatorial properties, including homology groups, homotopy groups, collapsibility, embeddability and spectral properties.

Babson, Hoffman and Kahle [1] considered $\pi_1(Y_2(n, p))$, the fundamental group of the random 2-complex, and showed that if $p < n^{-\alpha}$ for some arbitrary $\alpha > 1/2$ then this group is nontrivial whp. On the other hand, they proved that $\pi_1(Y_2(n, p))$ is trivial for $p > \sqrt{4 \log n / n}$, which means that the threshold probability for being simply connected should be close to $n^{-1/2}$. As a corollary of the first part of Theorem 1.1, we improve the upper bound on the threshold probability by a $\sqrt{\log n}$ factor.

Corollary 1.2 *Let $p = \frac{c}{\sqrt{n}}$ for some constant $c > \frac{1}{2}$. Then $Y_2(n, p)$ is simply connected whp.*

Proof. Theorem 1.1 shows that if H is the hypergraph corresponding to the 2-dimensional faces of $Y_2(n, p)$ then the triadic process propagates. Take the subcomplex C of $Y_2(n, p)$ containing the triangles used by the process to extend the edge set of the graph. This complex is collapsible to the star, so it has trivial fundamental group. Hence $\pi_1(Y_2(n, p))$ is also trivial. $\qquad\square$

1.1 Proof outline

Instead of exposing all the triples at once, we will be sampling them on the fly, trying to extend the edge set of the graph. At any point in the process, we say that a vertex triple $\{u, v, w\}$ is *open* if it spans exactly two edges but has not yet been sampled. We will also use the notation uvw for an open triple with edges uv and vw. By an *open triple at u*, we mean a triple uvw, i.e. one that has its missing edge adjacent to the vertex u.

Both the proofs of the upper bound and the lower bound consist of two phases. In the first phase we make one step at a time: we choose an open triple uniformly at random and expose it. With probability p the triple is selected, hence we can close it by adding the missing edge to the graph. The second phase proceeds in rounds: we simultaneously expose all the open triples and extend the edge set according to the outcome.

The essence of the proof is to track the behavior of certain variables throughout the process. As it turns out, this is not a very hard task to do in the second phase, using standard measure concentration inequalities. However, during the initial phase of the process, the codegrees (one of the variables we track) are not concentrated, which forces us to do a more careful analysis of the beginning of the process. For this we will use the differential equation method.

2 The differential equation method

The general idea of the differential equation method is the following. In order to say something about a discrete random process, we intend to track certain variables. We do so by translating recurrence relations (or difference equations) defining the one-step change in our variables into their continuous analogs, differential equations. Then we show that the variables follow the trajectories of the solutions of these differential equations. We analyze the process using some ideas from [2].

For simplicity, let us denote the graph we obtain after i samples by G_i. We consider the following random variables: $D_v(i)$ is the degree of the vertex v in G_i. $F_v(i)$ is the number of open triples at v, so it is the number of ways for v to gain a new incident edge in G_{i+1}. $X_{u,v}(i)$ is the codegree of u and v, i.e. the number of common neighbors of u and v in G_i.

To provide some insight, we first heuristically describe the process. Let us assume for now that the $D_v(i)$ are concentrated around some value $D(i)$, and similarly the $F_v(i)$ are approximately equal to some value $F(i)$. (For convenience, we drop the center of the star from consideration.)

In step $i + 1$ we sample an open triple uniformly at random, which corresponds to choosing a single such triple independently with probability $\frac{2}{\sum_v F_v(i)} \approx \frac{2}{nF(i)}$ because $\sum_v F_v(i)$ counts each open triple twice. As the number of open triples at a vertex v is about $F(i)$, the change in the degree of v we expect to see is

$$D(i+1) - D(i) \approx \frac{2p}{n}.$$

We gain open triples at v either if we successfully sample one of them (adding the edge vw), in which case new open triples are formed with the neighbors of w, or if we successfully sample a triple at some neighbor of v. On the other hand, we lose the sampled triple regardless of the outcome. Assuming all the codegrees are negligible compared to $D(i)$, this gives us

$$F(i+1) - F(i) \approx \frac{2}{n}(2pD(i) - 1).$$

To smooth out this discrete process, we introduce a continuous variable t and say that step i corresponds to time $t = \frac{i}{n^2}$. Let us also rescale D and F by considering the smooth functions d and f, where we want $d(t)$ to be approximately $D(i)/\sqrt{n}$ and $f(t)$ to be approximately $F(i)/n$. Then an easy calculation shows that $d'(t) \approx 2c$ and $f'(t) \approx 4cd(t) - 2$.

So far we assumed that the codegrees are negligible compared to the degrees, but since they are not concentrated, proving this still needs some thought. To this end, we introduce two more random variables. $Y_{u,v}(i)$ denotes the number of *open 3-walks* $uww'v$ from u to v, i.e. 3-walks where we require that uww' be open (but allowing $w = v$), and $Z_{u,v}(i)$ is the number of *open 4-walks* $uww'w''v$ (again, allowing vertex repetitions), where both uww' and $w'w''v$ are open. Note that $Y_{u,v}$ is *not* symmetric in u and v.

The point is that $Y_{u,v}$ and $Z_{u,v}$ are concentrated, and their one-step behavior can be described with fairly simple formulas. Indeed, analogously assuming

concentration around $Y(i)$, and $Z(i)$, and defining the rescaled functions y and z such that $y(t)$ is close to $Y(i)/\sqrt{n}$, and $z(t)$ is close to $Z(i)/n$, we obtain $y'(t) \approx \frac{2}{f(t)}\big((2cd(t)-1)y(t)+cz(t)\big)$ and $z'(t) \approx \frac{4}{f(t)}\big(2cy(t)f(t)-z(t)\big)$.

This illustrates why it is plausible to believe that the variables follow the tracks of the functions defined by the differential equations above. For the star $D(0) = 1$, $F(0) = n-2$, $Y(0) = 0$ and $Z(0) = n-3$, giving the initial conditions $d(0) = 0$, $f(0) = 1$, $y(0) = 0$ and $z(0) = 1$. The corresponding solution of our system of differential equations is

$$d(t) = 2ct \qquad\qquad f(t) = 1 - 2t + 4c^2t^2$$
$$y(t) = d(t)f(t) \qquad\qquad z(t) = f^2(t).$$

Next, we need to prove that our variables follow the prescribed trajectories up to some time T. Note that if $c \leq 1/2$ then f vanishes at $T_0 = \frac{1-\sqrt{1-4c^2}}{4c^2}$, so we expect the process to die around time T_0. In this case we choose T to be slightly less than T_0, when $c > 1/2$ we choose $T = \sqrt{\log n}$.

The allowed deviation of each variable will be defined by one of the error functions

$$g_1(t) = e^{Kt}n^{-1/6} \qquad\qquad \text{and} \qquad\qquad g_2(t) = (1+d(t))e^{Kt}n^{-1/6},$$

where K is some large constant depending on c.

Theorem 2.1 *Let $T \leq \sqrt{\log n}$ and K are defined as above. Then the following bounds hold with high probability for all vertices u and v and for every $i = 1, \ldots, T \cdot n^2$.*

$$D_v(i) \in \big(d(t) \pm g_1(t)\big)\sqrt{n} \qquad\qquad Y_{u,v}(i) \in \big(y(t) \pm g_2(t)\big)\sqrt{n}$$
$$F_v(i) \in \big(f(t) \pm g_1(t)\big)n \qquad\qquad Z_{u,v}(i) \in \big(z(t) \pm g_2(t)\big)n$$
$$X_{u,v}(i) \leq 50 \log n.$$

Proof sketch. We aim to show the concentration of each variable in each step separately. So let R represent any of the above variables, then to show $R(j)$ is in the prescribed interval, we condition on the event that concentration holds for every variable in all steps $i = 0, \ldots, j-1$.

Writing $R(j) = R(0) + \sum_{i=0}^{j-1}\big(R(i+1) - R(i)\big)$, we can use the difference equations described above to each term in the sum. This means that $R(j)$ is the sum of random variables restricted to intervals, allowing us to use some martingale concentration inequalities to bound the error probability. $\qquad\square$

3   Lower bound

To show that the process does not propagate when $c < 1/2$, we run the first phase up to time very close to T_0 (recall that T_0 is the first vanishing point of f). At this point all the degrees are less than $\sqrt{n}$. Now we start exposing open triples in rounds. Using an inductive argument, we can prove that all the degrees stay below $\sqrt{n}$ for at least $\Omega(\log n)$ rounds. Meanwhile, a parallel argument shows that as long as the degrees are bounded by $\sqrt{n}$, the expected total number of open triples drops exponentially in the rounds. Hence after $O(\log n)$ rounds we expect to see $o(1)$ open triples, meaning (by Markov's inequality) that the process dies whp while the degrees are still less than $\sqrt{n}$.

4   Upper bound

To show that the process propagates whp when $c > 1/2$, we run the first phase until time $T = \sqrt{\log n}$ and sample triples in rounds afterwards. Our plan is to give a sequence of lower bounds on the codegrees that increases exponentially in the number of rounds passed. At the end of the first phase we have no lower bound on X, but the fact that any open 4-walk contributes one with probability p^2 allows us to give a $\Omega(\log^2 n)$ lower bound already after the first round.

Note also that the vast majority of these codegrees are new, i.e. they correspond to open triples. Each such triple creates an edges with probability p, so we expect to see $\Omega(\sqrt{n} \log^2 n)$ new edges at each vertex, and consequently an increase of $\Omega(\log^4 n)$ in all the codegrees.

This interaction between the degrees and codegrees, together with some Chernoff-type concentration arguments allow us to prove that after i rounds all the codegrees are bounded from below by $(\varepsilon \log n)^{2^i}$ for some small ε (as long as this value is below n, of course). Hence we reach the complete graph in $O(\log \log n)$ rounds, in other words the process propagates.

References

[1] Babson, E., C. Hoffman and M. Kahle, *The fundamental group of random 2-complexes*, Journal of the American Mathematical Society **24** (2011), 1–28.

[2] Bohman, T., *The Triangle-Free Process*, Advances in Mathematics **221** (2009), 1653–1677.

[3] Bohman, T., A. Frieze and E. Lubetzky, *Random triangle removal*, Submitted.

[4] Coja-Oghlan, A., M. Onsjö and O. Watanabe, *Propagation Connectivity of Random Hypergraphs*, The Electronic Journal of Combinatorics **19** (2012) P17, 25pp.

[5] Easley, D. and J. Kleinberg, "Networks, crowds, and markets: reasoning about a highly connected world", Cambridge University Press, 2010.

[6] Erdős, P., S. Suen and P. Winkler, *On the size of a random maximal graph*, Random Structures & Algorithms **6** (1995), 309–318.

[7] Linial, N. and R. Meshulam, *Homological connectivity of random 2-dimensional complexes*, Combinatorica **26** (2006), 475–487.

[8] Wormald, N., *The differential equation method for random graph processes and greedy algorithms*, "Lectures on Approximation and Randomized Algorithms", PWN, Warsaw, 1999, 73–155.