Supplementary material to Learning multifractal structure in large networks

Austin R. Benson, Carlos Riquelme, Sven P. Schmit
Institute for Computational and Mathematical Engineering
Stanford University
{arbenson, rikel, schmit} @ stanford.edu

February 21, 2014

1 Proof of Corollary 4.

1.1 Expected number of triangles

Let u, v and w be three random nodes of the graph. We define $E_{u,v,w}$ to be the event that there is a triangle between u, v, w and, similarly, let $E^{(i)}_{u,v,w}$ be the event that there is a triangle between u, v, w in H_i, $i = 1, \ldots, k$. Then, by Theorem 2,

$$\mathbb{P}(E_{u,v,w}) = \prod_{i=1}^{k} \mathbb{P}(E^{(i)}_{u,v,w}) = \mathbb{P}(E^{(1)}_{u,v,w})^k. \quad (1)$$

Now, we compute the probability of a triangle happening between three random nodes according to $\mathcal{W}_1$:

$$\begin{aligned}
\mathbb{P}(E^{(1)}_{u,v,w}) &= \mathbb{P}((u,v),(u,w),(v,w) \in E) \\
&= \sum_{i,j,t \in \mathcal{C}} \mathbb{P}((u,v),(u,w),(v,w) \in E | c_1^u = i, c_1^v = j, c_1^w = t)\ \mathbb{P}(c_1^u = i, c_1^v = j, c_1^w = t) \\
&= \sum_{i,j,t \in \mathcal{C}} \mathbb{P}((u,v),(u,w),(v,w) \in E | c_1^u = i, c_1^v = j, c_1^w = t)\ \mathbb{P}(c_1^u = i)\mathbb{P}(c_1^v = j)\mathbb{P}(c_1^w = t) \\
&= \sum_{i,j,t \in \mathcal{C}} \mathbb{P}((u,v) \in E | c_1^u = i, c_1^v = j)\mathbb{P}((u,w) \in E | c_1^u = i, c_1^w = t)\mathbb{P}((v,w) \in E | c_1^v = j, c_1^w = t)\ l_i l_j l_t \\
&= \sum_{i,j,t \in \mathcal{C}} p_{ij} p_{it} p_{jt}\ l_i l_j l_t =: s_3.
\end{aligned}$$

By (1), we conclude that

$$\mathbb{P}(E_{u,v,w}) = s_3^k. \quad (2)$$

We can now compute the expected number of triangles C_3:

$$\mathbb{E}[C_3] = \sum_{\substack{S \subset V \\ |S|=3}} \mathbb{1}(E_S) = \binom{n}{3} \mathbb{P}(E_{u,v,w}) = \binom{n}{3} s_3^k. \quad (3)$$

1.2 Expected number of t-cliques

We can generalize the proof for the expected number of triangles to the expected number of t-cliques. Consider t random nodes $S = \{u_1, \ldots, u_t\}$ and let E_S be the event that they form a t-clique. Furthermore, let $E_S^{(i)}$ be the event that they form a t-clique in H_i, $i = 1, \ldots, k$. Then by Theorem 2,

$$\mathbb{P}(E_S) = \prod_{i=1}^{k} \mathbb{P}(E_S^{(i)}) = \mathbb{P}(E_S^{(1)})^k.$$

Also, it follows that

$$\mathbb{P}(E_S^{(1)}) = \sum_{i_1, \ldots, i_t \in \mathcal{C}} \left(\prod_{\substack{j,q \in [t] \\ j \neq q}} p_{i_j i_q} \right) l_{i_1} l_{i_2} \cdots l_{i_t} := s_t$$

Finally, let C_t be the expected number of $t-$cliques in G. We conclude that

$$\mathbb{E}[C_t] = \binom{n}{t} s_t^k. \tag{4}$$

1.3 Expected number of wedges

Let u, v, w be three distinct nodes of G. We define A to be the event that there is a wedge centered at u in G, that is, $A = \{(u, v), (u, w) \in E(G)\}$. Similarly, as in previous sections, we define $A^{(i)}$ to be the event restricted to H_i. By Theorem 2,

$$\mathbb{P}(A) = \prod_{i=1}^{k} \mathbb{P}(A^{(i)}) = \mathbb{P}(A^{(1)})^k.$$

Now, by considering only H_1,

$$\begin{aligned}
\mathbb{P}(A^{(1)}) = \mathbb{P}((u, v), (u, w) \in E) &= \sum_{i,j,t \in \mathcal{C}} \mathbb{P}((u, v), (u, w) \in E | c_1^u = i, c_1^v = j, c_1^w = t) \mathbb{P}(c_1^u = i, c_1^v = j, c_1^w = t) \\
&= \sum_{i,j,t \in \mathcal{C}} \mathbb{P}((u, v) \in E | c_1^u = i, c_1^v = j) \mathbb{P}((u, w) \in E | c_1^u = i, c_1^w = t)\ l_i l_j l_t \\
&= \sum_{i,j,t \in \mathcal{C}} p_{ij} p_{it}\ l_i l_j l_t =: \omega.
\end{aligned}$$

It follows that the expected number of wedges S_2 in G is given by

$$\mathbb{E}[S_2] = n \binom{n-1}{2} \omega^k. \tag{5}$$

1.4 Variance of the number of edges

Let X_{ij} be the indicator random variable of the event $(v_i, v_j) \in E$ for $i \neq j$. We also define $X = \sum_{i<j} X_{ij}$, the total number of edges. We compute the second moment of X as follows:

$$\begin{aligned}
\mathbb{E}[X^2] &= \mathbb{E}\left[\left(\sum_{i<j} X_{ij}\right)\left(\sum_{i<j} X_{ij}\right)\right] \\
&= \mathbb{E}\left[\sum_{i<j} X_{ij}^2\right] + \mathbb{E}\left[\sum_{i,j\neq k, i<j,k} X_{ij}X_{ik}\right] + \mathbb{E}\left[\sum_{i\neq k, j\neq z, i<j, k<z} X_{ij}X_{kz}\right] \\
&= \mathbb{E}[X] + 2\mathbb{E}[S_2] + \sum_{i\neq k, j\neq z, i<j, k<z} \mathbb{E}\left[X_{ij}\right]\mathbb{E}\left[X_{kz}\right] \\
&= \mathbb{E}[X] + 2\mathbb{E}[S_2] + \binom{n}{2}\binom{n-2}{2}\mathbb{P}\left((v_i, v_j) \in E\right)^2 \\
&= \mathbb{E}[X] + 2\mathbb{E}[S_2] + \binom{n}{2}\binom{n-2}{2}s^{2k}.
\end{aligned}$$

Hence, we see that

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \mathbb{E}[X] + 2\mathbb{E}[S_2] + \binom{n}{2}\binom{n-2}{2}s^{2k} - \mathbb{E}[X]^2 \\
&= \mathbb{E}[X](1 - \mathbb{E}[X]) + 2n\binom{n-1}{2}\omega^k + \binom{n}{2}\binom{n-2}{2}s^{2k} \\
&= \binom{n}{2}s^k\left(1 - \binom{n}{2}s^k\right) + 2n\binom{n-1}{2}\omega^k + \binom{n}{2}\binom{n-2}{2}s^{2k}.
\end{aligned}$$

1.5 d-stars and degree distribution

A $d-$star centered at node u is a graph containing $d+1$ vertices whose edges go from u to each of the other d vertices in the graph. For example, a wedge is a 2-star. We start by noting the following key and simple fact: the number of vertices with degree d in a graph G equals the number of copies of $d-$stars in G that are not part of any $(d+1)-$star in G. Let us define X_d to be the random variable that counts the number of $d-$stars in G, for any $d \in [n-1]$.

Let $d' > d$ and suppose that vertex u has degree d'. Then, u will contribute with $\binom{d'}{d}$ stars to X_d.

We define V_d to be the random variable that counts the number of nodes with degree $\geq d$. Similarly, we denote by E_d the number of nodes with degree d. We see that

$$E_d = V_d - V_{d+1},$$

which directly implies

$$\mathbb{E}[E_d] = \mathbb{E}[V_d] - \mathbb{E}[V_{d+1}].$$

Our goal is to write V_d as a function of X_d and X_{d+1}. We see that $E_{n-1} = X_{n-1}$ and

$$E_d = X_d - \sum_{i=d+1}^{n}\binom{i}{d}E_i.$$

Taking expectations, we conclude that

$$\mathbb{E}[E_d] = \mathbb{E}[X_d] - \sum_{i=d+1}^{n} \binom{i}{d} \mathbb{E}[E_i].$$

We can compute the expected number of $d-$stars in G. Let $u_1, \ldots, u_{d+1}$ be $d+1$ distinct nodes, and let S be the event that there is a $d-$star centered at u_1 in G. In other words, $S = \{(u_1, u_2), (u_1, u_3), \ldots, (u_1, u_{d+1}) \in E\}$. Let $S^{(i)}$ be the event restricted to H_i. By Theorem 2,

$$\mathbb{P}(S) = \prod_{i=1}^{k} \mathbb{P}(S^{(i)}) = \mathbb{P}(S^{(1)})^k.$$

Now, let us compute $\mathbb{P}(S^{(1)})$:

$$\begin{aligned}
\mathbb{P}(S_1) &= \mathbb{P}((u_1, u_2), (u_1, u_3), \ldots, (u_1, u_{d+1}) \in E) \\
&= \sum_{i_1, \ldots, i_{d+1} \in \mathcal{C}} \mathbb{P}((u_1, u_2), (u_1, u_3), \ldots, (u_1, u_{d+1}) \in E | c_1^{u_j} = i_j \forall j) \; \mathbb{P}(c_1^{u_j} = i_j \forall j) \\
&= \sum_{i_1, \ldots, i_{d+1} \in \mathcal{C}} \mathbb{P}((u_1, u_2) \in E | c_1^{u_1} = i_1, c_1^{u_2} = i_2) \cdots \mathbb{P}((u_1, u_{d+1}) \in E | c_1^{u_1} = i_1, c_1^{u_{d+1}} = i_{d+1}) \; \prod_{j=1}^{d+1} l_{i_j} \\
&= \sum_{i_1, \ldots, i_{d+1} \in \mathcal{C}} \left(\prod_{j=2}^{d+1} p_{i_1 i_j} \right) \; \prod_{j=1}^{d+1} l_{i_j}.
\end{aligned}$$

We can now compute the expected number of $d-$stars in G:

$$\begin{aligned}
\mathbb{E}[X_d] &= n \binom{n-1}{d} \mathbb{P}(S) = n \binom{n-1}{d} \mathbb{P}(S_1)^k \\
&= n \binom{n-1}{d} \; \left[\sum_{i_1, \ldots, i_{d+1} \in \mathcal{C}} \left(\prod_{j=2}^{d+1} p_{i_1 i_j} \right) \; \prod_{j=1}^{d+1} l_{i_j} \right]^k.
\end{aligned}$$

2 Recovered measures for real-world networks

In this section, we provide the recovered measures found by the method of moments algorithm for the real data sets.

2.1 Gnutella

For $m = 2$, k is 16, and

$$\mathbf{P} = \begin{pmatrix} 0.9424 & 0.2241 \\ 0.2241 & 0.7232 \end{pmatrix}, \quad \ell = \begin{pmatrix} 0.5869 \\ 0.4311 \end{pmatrix}.$$

For $m = 3$, k is 11, and

$$\mathbf{P} = \begin{pmatrix} 0.9562 & 0.0873 & 0.2602 \\ 0.0873 & 0.6078 & 0.1486 \\ 0.2602 & 0.1486 & 0.7090 \end{pmatrix}, \quad \ell = \begin{pmatrix} 0.4249 \\ 0.1869 \\ 0.3882 \end{pmatrix}.$$

2.2 Citation

For $m = 2$, k is 16, and

$$\mathbf{P} = \begin{pmatrix} 1.0000 & 0.0567 \\ 0.0567 & 0.9202 \end{pmatrix}, \quad \ell = \begin{pmatrix} 0.2219 \\ 0.7781 \end{pmatrix}.$$

For $m = 3$, k is 10, and

$$\mathbf{P} = \begin{pmatrix} 0.9999 & 0.7387 & 0.9930 \\ 0.7387 & 0.7072 & 0.0062 \\ 0.9930 & 0.0062 & 0.9003 \end{pmatrix}, \quad \ell = \begin{pmatrix} 0.0487 \\ 0.3794 \\ 0.5719 \end{pmatrix}.$$

2.3 Facebook

For $m = 2$, k is 12, and

$$\mathbf{P} = \begin{pmatrix} 1.0000 & 0.0653 \\ 0.0653 & 0.9679 \end{pmatrix}, \quad \ell = \begin{pmatrix} 0.1969 \\ 0.8031 \end{pmatrix}.$$

For $m = 3$, k is 8, and

$$\mathbf{P} = \begin{pmatrix} 1.0 & 0 & 1.0 \\ 0 & 0.7204 & 1.0 \\ 1.0 & 1.0 & 0.9696 \end{pmatrix}, \quad \ell = \begin{pmatrix} 0.5933 \\ 0.3373 \\ 0.0694 \end{pmatrix}.$$

2.4 Twitter

For $m = 2$, k is 17, and

$$\mathbf{P} = \begin{pmatrix} 0.5312 & 0.1047 \\ 0.1047 & 0.9358 \end{pmatrix}, \quad \ell = \begin{pmatrix} 0.2194 \\ 0.7806 \end{pmatrix}.$$

For $m = 3$, k is 11, and

$$\mathbf{P} = \begin{pmatrix} 0.5132 & 1.0000 & 0 \\ 1.0000 & 1.0000 & 1.0000 \\ 0 & 1.0000 & 0.9311 \end{pmatrix}, \quad \ell = \begin{pmatrix} 0.3648 \\ 0.0598 \\ 0.5754 \end{pmatrix}.$$