Random Digraphs : Some Concentration Results

Joint work with Kunal Dutta and Joel Spencer.

ICM-2010 Satellite Conference on Algebraic and Probabilistic Aspects of Combinatorics and Computing,
Indian Institute of Science, Bangalore,

August 29- September 3, 2010.

C R Subramanian

The Institute of Mathematical Sciences, Chennai.

Random Graph models

- $V = \{1, 2, \ldots, n\}$. $G = (V, E)$. $p = p(n)$.
- $G \in \mathcal{G}(n, p)$: $e \in E$ independently with probability p.
- $D \in \mathcal{D}(n, p)$: $p \leq 0.5$. Choose $G \in \mathcal{G}(n, 2p)$. Orient each $e \in E$ uniformly and independently.
- $D \in \mathcal{D}_2(n, p)$: $p \leq 0.5$. Choose each $e \in V \times V - \{(u, u)\}_u$ independently with probability p. Allows 2-cycles.

$\alpha(G)$ and $\omega(G)$

- $G \in \mathcal{G}(n, p)$, $p \leq 0.5$.
- $\omega(G) =$ maximum size of a clique in G.
- $\alpha(G) =$ maximum size of an indep set in G.
- Determination of $\omega(G)$ and $alpha(G)$ are equivalent.
- $\omega(G \in \mathcal{G}(n, p))$ and $\alpha(G \in \mathcal{G}(n, 1-p))$ have the same distribution.
- $\mathbf{Pr}(\omega(G \in \mathcal{G}(n, p)) = b) = \mathbf{Pr}(\alpha(G \in \mathcal{G}(n, 1-p)) = b)$.

$\alpha(G)$ and $\omega(G)$

- concentration of $\omega(G)$:
- $\omega(G)$ is tightly concentrated in just two values.
- Eg : $p = 1/2 \Rightarrow \omega \in \{k, k+1\}$ almost surely
- for some $k = 2\log n - 2\log\log n + O(1)$.
- No simple closed-form expression for k.
-
- Concentration of $\alpha(G)$:
- Assume $p \geq C/n$. $q = (1-p)^{-1}$. Almost surely,
- $\alpha(G) = \frac{2}{\ln q}(\ln np - \ln\ln np \pm O(1))$.
- α is not tightly concentrated.

$mat(D)$ and $mas(D)$

- Similar phenomena in random directed graphs.
- $D \in \mathcal{D}(n, p)$. $p \leq 0.5$.
- $mat(D)$ = maximum size induced acyclic tournament in D.
- $mas(D)$ = maximum size induced acyclic subgraph in D.
- $mat(D)$ is 2-poin
- $mas(D)$ = maximum size induced acyclic tournament in D.t concentrated or even one-point concentrated. Also, admits sharp thresholds.
- Unlike $\omega(G)$, admits a nice closed form expression.
- $mas(D)$ has coarse concentration like $\alpha(G)$.

$\omega(G)$ vs $mat(D)$ and $\alpha(G)$ vs $mas(D)$

- $D \in \mathcal{D}(n,p)$ and $G \in \mathcal{G}(n,p)$; $b \geq 1$.
- $\mathbf{Pr}[mas(D) \geq b] \geq \mathbf{Pr}[\alpha(G) \geq b]$.
- τ - a fixed linear ordering of V.
- $\mathbf{Pr}(mas(D) \geq b)$ is at least the probability that $D[A]$ is consistent with τ for some $A, |A| = b$.
- Equals $\mathbf{Pr}(\omega(G) \geq b)$.
-
- similarly, for $mat(D)$,
- $\mathbf{Pr}[mas(D) \geq b] \geq \mathbf{Pr}[\alpha(G) \geq b]$.

2-point concentration of $mat(D)$

(Kunal and CRS)

- $D \in \mathcal{D}(n, p)$, $p \geq 1/n$.
- $b^* = \lfloor 2(\log_{p^{-1}} n) + 0.5 \rfloor$.
- almost surely, $mat(D) \in \{b^*, b^* + 1\}$.
- Fact : A *dag* has at most one directed hamilton path.
-
- **Proof Sketch :** For $b \geq 1$, define
- X_b = number of induced acyclic tournaments of size b.
- $E[X_b] = \binom{n}{b} b! p^{\binom{b}{2}} \approx \left(np^{(b-1)/2}\right)^b$.
- $E[X_b] \to 0$ for $b = b^* + 2$.
- Hence $mat(D) \leq b^* + 1$ almost surely.

2-point concentration of $mat(D)$

- To prove $mat(D) \geq b^*$ almost surely,
- Show : $\mu = E[X_b^*] \to \infty$ and also
- $\mathbf{Pr}(X_{b^*} = 0) \leq \mathbf{Pr}(|X_{b^*} - \mu| \geq \mu) \to 0$ using Chebyshev.
- Suffices to show that, for $b = b^*$,
- $Var(X_b) \leq \mu + \mu \left(\sum_{i,j:|A_i \cap A_j| \in [2,b-1]} E(X_j|X_i) \right)$.
- $Var(X) \leq \mu + o(\mu^2)$.

One point concentration of $mat(D)$

- $D \in \mathcal{D}(n,p)$, $w = w(n) \to \infty$ sufficiently slowly.
- $d = 2\log_{p^1} n + 1$ and $\delta = \lceil d \rceil - d$.
- Suppose $\frac{w}{\ln n} \leq \delta \leq 1 - \frac{w}{\ln n}$ for large values of n.
- almost surely, $mat(D) = \lfloor d \rfloor$.
- $\delta \leq 0.5 \Rightarrow \lfloor d \rfloor = b^*$.
- $\delta > 0.5 \Rightarrow \lfloor d \rfloor = b^* + 1$.

one-point concentration

- p fixed but arbitrary.
- $mat(D)$ is one-point concentrated for each n from a subset of integers of density 1.
- **Proof sketch :**
- Every n must be of the form $t^{(k-1-\delta)/2}$ for some $k \geq 0$. $t = p^{-1}$.
- every good n should satisfy
- $t^{\frac{k-1-\delta}{2}+\frac{w}{2\ln n}} \leq n \leq t^{\frac{k-1-\delta}{2}-\frac{2}{2\ln n}}$.
- does not hold when p varies with n. Eg : $p = n^{-2/3}$.

threshold phenomena and algorithms

- For every i, there exist $p_i = p_i(n)$ and $q_i = q_i(n)$ with
- $q_i = o(p_i)$ such that almost surely
- $p \geq p_i + q_i \Rightarrow mat(D) \geq i$
- $p \leq p_i - q_i \Rightarrow mat(D) < i$.
- sharp threshold exists.
-
- $lb_i(n) = n^{-4/(2i-1-2w/\ln n)}$ and $ub_i(n) = n^{-4/(2i-1+\frac{2w}{\ln n})}$
- $p_i(n) = (lb_i(n) + ub_i(n))/2$,
- $q_i(n) = (ub_i(n) - lb_i(n))/2$.

improved algorithm

- $w = w(n) \to \infty$. almost surely,
- every maximal solution is of size at least
- $d = \lfloor \delta \log_{p^{-1}} n \rfloor$ where $\delta = 1 - \frac{\ln(\ln n + w)}{\ln n}$.
- $c \geq 1$ constant. $p \geq n^{-1/c^2}$.
- $\exists$ deter. poly time algor A which almost surely
- finds a solution of size at least $\log_{p^{-1}} n + c\sqrt{\log_{p^{-1}} n}$.
-
- Can one find in poly time a soln of size at least
- $(1+\epsilon)\log_{p^{-1}} n$, for some fixed $\epsilon > 0$.

Results on $mas(D)$

- $D \in \mathcal{D}(n, p)$, $p \leq 0.5$.
- difficulty : Given A, what is
- $\mathbf{Pr}(D[A]$ is acyclic) ?
-
- $|mas(D) - mas(D')| \leq 1$ if D and D' differ only with respect to a single vertex.
- Using a vertex-exposure martingale and Azuma's martingale inequality, with $\mu = E[mas(D)]$,
- $|mas(D) - \mu| \leq w\sqrt{n}$ for any $w \to \infty$.
- the "likely" values of $mas(D)$ still not known.

Some easy consequences (CRS)

- $D \in \mathcal{D}(n, p)$, $p \leq 0.5$. Define $q = (1 - p)^{-1}$, $w = np$.
- $mas(D) \leq \lfloor 2 \log_q n + 1 \rfloor$.
- $mas(D) \geq \frac{2}{\ln q} (\ln w - \ln \ln w - O(1))$.
- the ratio of the two bounds can be very large,
- particularly, if $p = n^{-1+o(1)}$.
-
- conj : Is it true that
- $mas(D) = \frac{2(\ln w)}{\ln q} (1 \pm o(1))$?

improved upper bound on $mas(D)$ (due to Spencer)

- Fix A of size b.
- $D[A]$ is acyclic only if $\exists A = A_1 \cup A_2$
- with no arc going from A_2 to A_1.
-
- $\mathbf{Pr}(D[A] \text{ is acyclic }) \leq 2^b (1-p)^{b^2/4}$.
- $\mathbf{Pr}(\exists A, |A| = b : D[A] \text{ is acyclic}) \leq \left(\frac{2en}{b}\right)^b (1-p)^{b^2/4}$.
- $mas(D) \leq \frac{4 \ln w}{\ln q}$ almost surely.
- the ratio of the bounds is now at most two.

improved bounds on $mas(D)$ (due to CRS)

- constant 4 can be brought down further.
- For suitable k, choose a k-partition instead of a bipartition.
- k cannot become too large. to be chosen carefully.
- choose $b = \lfloor \frac{2}{\ln q}(\ln w + 3e) \rfloor$ and
- choose k the integer nearest to $2(\ln w)(3e)^{-1} + 2$.
- $mas(D) \leq b$ almost surely.
-
- $mas(D) = \frac{2(\ln w)}{\ln q}(1 \pm o(1))$ almost surely.

additively improved bounds on $mas(D)$ (Kunal and CRS)

- the ratio of the bounds is $1 + o(1)$.
- Still, an additive gap of $\frac{\ln \ln w}{\ln q}$ exists.
- $Y = Y(b) = |\{(A, \sigma) : |A| = b, \sigma \text{ certifies } A\}|$.
- $Y = \sum_{i \leq m} Y_i$ where $m = (n)_b$.
- $(A_1, \sigma_1), \ldots, (A_m, \sigma_m)$.
- $E[Y] = (n)_b (1 - p)^{\binom{b}{2}}$.
- $b^* = \lfloor \frac{2 \ln w}{\ln q} - X \rfloor$ where
- $X = W$ if $p \geq n^{-1/3+\epsilon}$
- $X = W/(\ln q)$ if $p \geq n^{-1/2} (\ln n)^2$.

additive improvements

- At $b = b^*$, $E[Y] \to \infty$ as $n \to \infty$.
- $Var(Y) \leq \mu + \mu^2 \cdot M$ where
- $M = \sum_{j:2 \leq |A_i \cap A_j| \leq b} E[Y_j | Y_i = 1]/\mu$.
- $E[Y_j | Y_i = 1] = (1-p)^{\binom{b}{2} - \binom{l}{2}} \left(\frac{1-2p}{1-p}\right)^{i(\pi)}$.
- here, $l = |A_i \cap A_j|$. π is the relative ordering of $A_i \cap A_j$ with respect to the ordering imposed by σ_i.
- Uses the following well-known fact :
- $\sum_{\sigma \in S_n} q^{i(\sigma)} = (1+q)(1+q+q^2)\dots(1+q+\dots+q^{n-1})$.

additive improvements

- $\exists W > 0 : \forall \epsilon > 0,\ p \geq n^{-1/3+\epsilon}$,
- $mas(D) \geq \frac{2(\ln w)}{\ln q} - W$ almost surely.
-
- $\exists W > 0 : \forall p \geq n^{-1/2}(\ln n)^2$,
- $mas(D) \geq \frac{2}{\ln q}(\ln w - W)$ almost surely.
-
- $p \geq n^{-1/2+\epsilon} \Rightarrow mas(D) \leq \frac{2}{\ln q}(\ln w + \ln(7e))$ almost surely.

Algorithms

- Every maximal induced dag is of size at least $\delta(\log_q w)$ for some $\delta \to 1$ as $w \to 1$.
-
- An induced dag of size at least $\log_q w + c\sqrt{\log_q w}$ can be found almost surely.
-
- Most of these results carry over to the $\mathcal{D}_2(n, p)$ model with some small changes.

Further work

- Further progress made in reducing the additive gap.
- A tighter concentration based on Talagrand's inequality is possible. Details later.

Thank You