Lecture 03: Concentration Inequalities II

Nikola Zlatanov

Innopolis University

Advanced Statistics

6-th of Feb to 13-th of Feb, 2023

Motivation

$E(X) = 1 \cdot p + 0(1-p) = p$

- Do we always get such fast Gaussian tails for other distributions?
- Of course not. In fact, we will play with the standard Bernoulli distribution and show that the its sum is very poorly bounded by the Hoeffding's inequality.
- The standard Bernoulli distribution, denoted in short by $\text{Ber}(p)$, is the following $\Pr\{X = 1\} = p$ and $\Pr\{X = 0\} = 1 - p$. Hence, here the standard Bernoulli RV X takes outcomes 0 and 1, instead of outcomes -1 and 1 as in the symmetric Bernoulli.
- For $X \sim \text{Ber}(p)$, then $E[X] = p$ and $\text{VAR}[X] = p(1 - p)$.
- We can use the sum of standard Bernoulli RVs, in the limit when $p \to 0$, to show that Hoeffding's' inequality does not work well on that tail and then derive a another inequality, Chernoff's inequality, which works well on that tail.

Poisson Distribution

- Let S_n be defined as

$$S_n = \sum_{i=1}^{n} X_i,$$

i.i.d.

where X_i is a standard Bernoulli RV with $\Pr\{X_i = 1\} = p = \mu/n$ and $\Pr\{X_i = 0\} = 1 - \mu/n$.

- In this case $E[S_n] = \mu$. Hence, $E[S_n]$ is constant independent of how big n is.
- Thm (Poisson Limit Theorem): Let $X_1, X_2, ..., X_n$ be i.i.d. according to $\text{Ber}(\mu/n)$. Then $S_n = \sum_{i=1}^{n} X_i$ converges in distribution to the Poisson distribution as $n \to \infty$.
- **Note**: The normalized sum of S_n, i.e., $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - E[X_i]}{\sigma}$, when $X_i \sim \text{Ber}(\mu/n)$, does not converge in distribution to Gaussian since the parameter $p = \mu/n$ in $\text{Ber}(\mu/n)$ is not fixed, but depends on n. If it were fixed, then the normalized sum of S_n would converge to Gaussian in distribution.

Poisson Distribution

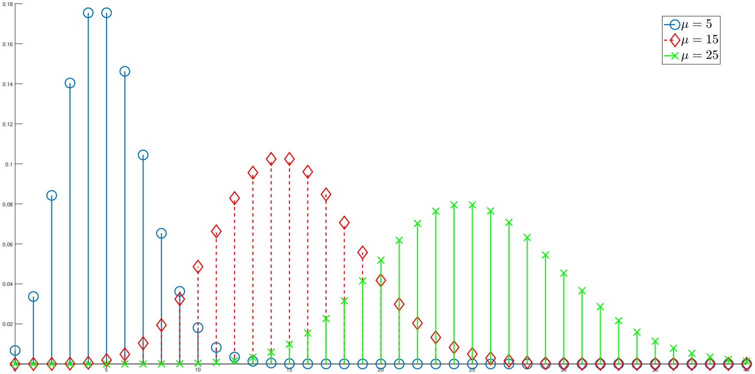
- V is said to be Poisson distributed with parameter μ, written shorthand as $V \sim \text{Pois}(\mu)$, where $V \in \{0, 1, 2, ...\}$, and it's distribution is

$$\Pr\{V = k\} = \frac{\mu^k}{k!} e^{-\mu}, \text{ for } k = 0, 1, 2, ...$$

  where $E[V] = \mu$ and $\text{VAR}[V] = \mu$ hold.
- The Poisson distribution is used for modelling rare events, such as the sum of Bernoulli where $p \to 0$ or $p \to 1$.
- The Poisson distribution is used for modelling rare disease among a population, number of car accidents in a day, number of emergency calls at a hospital in a day, etc.

Poisson Distribution

- Graph of the distribution



Poisson Tail

- If $V \sim \text{Pois}(\mu)$, then the tail of V is $c_1 e^{-c_2 t^2}$

$$\Pr\{V \geq t\} = \sum_{k=t}^{\infty} \frac{\mu^k}{k!} e^{-\mu},$$

which is a difficult expression to derive.

- Let's try a different method via the sum of n standard Bernoulli RVs with $\text{Ber}(\mu/n)$ for obtaining the tail.
- Thm (Chernoff's inequality): Let $X_i \sim \text{Ber}(p)$ be independent RVs, with $p = \mu/n$. Let $S_n = \sum_{i=1}^{n} X_i$, where $E[S_n] = \sum_{i=1}^{n} E[X_i] = \mu$. Then

$$\Pr\{S_n \geq t\} \leq e^{-\mu} \left(\frac{e\mu}{t}\right)^t, \forall t \geq \mu$$

Poisson Tail

- Proof (via MGF): Let $\lambda \geq 0$, which will be determined later, then

$$\begin{aligned}\Pr\{S_n \geq t\} &\overset{(a)}{=} \Pr\{e^{\lambda S_n} \geq e^{\lambda t}\} \overset{(b)}{\leq} \frac{E[e^{\lambda S_n}]}{e^{\lambda t}} = \frac{E\left[e^{\lambda \sum_{i=1}^{n} X_i}\right]}{e^{\lambda t}} \\ &= \frac{E\left[\prod_{i=1}^{n} e^{\lambda X_i}\right]}{e^{\lambda t}} \overset{(c)}{=} \frac{\prod_{i=1}^{n} E\left[e^{\lambda X_i}\right]}{e^{\lambda t}} \end{aligned} \tag{1}$$

where (a) is due to multiplying by $\lambda \geq 0$ and exponentiation from both sides, and (b) is due to Markov's inequality, and (c) is due to the independence of the X_i.
On the other hand,

$$E\left[e^{\lambda X_i}\right] = pe^{\lambda \times 1} + (1-p)e^{\lambda \times 0} = pe^{\lambda} + (1-p) = 1 + p(e^{\lambda} - 1) \tag{2}$$

Poisson Tail

If we insert (2) into (1), it will be very hard to work with $\prod_{i=1}^{n}(1+p(e^{\lambda}-1)) = (1+p(e^{\lambda})-1)^n$.

That is why, we would like to upper bound $1+p(e^{\lambda}-1)$ by one exponential since then exponential on power n is easy to work with.

To this end, we use $1+y \leq e^y$, which after setting $y = p(e^{\lambda}-1)$ leads to

$$1+p(e^{\lambda}-1) \leq \exp(p(e^{\lambda}-1)) \tag{3}$$

Poisson Tail

Inserting (3) into (2), and then (2) into (1), we obtain

$$\Pr\{S_n \geq t\} \leq \frac{\prod_{i=1}^{n} \exp(p(e^\lambda - 1))}{e^{\lambda t}} = \frac{\exp((e^\lambda - 1)np)}{e^{\lambda t}}$$
$$\stackrel{(a)}{=} e^{-\lambda t} \exp((e^\lambda - 1)\mu) = \exp((e^\lambda - 1)\mu - \lambda t), \qquad (4)$$

where (a) comes from our initial assumption that $p = \mu/n$.

Now, to make the bound more tight, we need to minimize the right hand side of (4), which means we need to minimize $(e^\lambda - 1)\mu - \lambda t$ for $\lambda \geq 0$. This leads to

$$\frac{\partial}{\partial \lambda}(e^\lambda - 1)\mu - \lambda t = e^\lambda \mu - t = 0,$$

from where $\lambda = \ln(t/\mu)$, which holds only for $t \geq \mu$ since λ must satisfy $\lambda \geq 0$.

$a \cdot \ln(b) = \ln(b^a)$

Inserting $\lambda = \ln(t/\mu)$ into (4), we obtain

$$\Pr\{S_n \geq t\} \leq \exp\left(\left(\frac{t}{\mu} - 1\right)\mu - \ln\left(\frac{t}{\mu}\right)t\right) = \exp\left((t - \mu + \ln\left(\frac{\mu^t}{t^t}\right)\right)$$

$$= e^{-\mu} e^t \frac{\mu^t}{t^t} = e^{-\mu}\left(\frac{e\mu}{t}\right)^t, \text{ for } t \geq \mu$$

Q.E.D.

Poisson Tail

$\hat{p} = \frac{\mu}{n}$ $E(S_n) = \mu$

- Now, the tail we obtained

$$\Pr\{S_n \geq t\} \leq e^{-\mu} \left(\frac{e\mu}{t}\right)^t, \text{ for } t \geq \mu \tag{5}$$

  does not depend on n.
- Since $V \sim \text{Pois}(\mu)$ is obtained as $\lim_{n\to\infty} S_n$, it turns out that the tail of $\Pr\{S_n \geq t\}$ also holds for $\Pr\{V \geq t\}$ for any n, since the tail in (5) does not depend on n.
- To summarize, (5) is both the tail of the Poisson distribution and the tail of the sum $S_n = \sum_{i=1}^n X_i$, where $X_i \sim \text{Ber}(\mu/n)$. Also, this tail is tight!

Poisson Tail

- Prove at home that if we use Hoeffding's inequality to obtain the tail of $S_n = \sum_{i=1}^{n} X_i$, where $X_i \sim \mathrm{Ber}(\mu/n)$, the tail will be very loose for large n.

$$\Pr\left\{ \frac{1}{\sqrt{n}} \sum \frac{X_i - \mu/n}{\frac{\mu}{n}\left(1 - \frac{\mu}{n}\right)} > t \right\} \le \exp\left(-t^2/2\sigma^2\right)$$

$n \to \infty$

$\sigma = 1$

$a = 0$

$$\Pr\left(\sum_{i=1}^{n} X_i > t\right) \le f_{u_n}(t) \le 1$$

$P \to f_{D_n}$

$p = \frac{\mu}{n}$

Simplification of the Poisson Tail

- Now, the Poisson tail that we obtained

$$\Pr\{S_n \geq t\} \leq e^{-\mu} \left(\frac{e\mu}{t}\right)^t, \text{ for } t \geq \mu$$

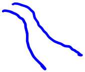
  is hard to understand intuitively of what is really going on. Let's try to simplify it, and understand it better, intuitively.
- To this end, we look separately at
  - $t \gg \mu$, which is called the "Large Deviation Regime"
  - $t > \mu$ but $t \approx \mu$, which is called the "Small Deviation Regime"

$t = \mu + \varepsilon \quad \varepsilon \to 0$

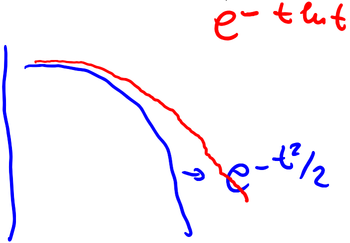
Large Deviation Regime of the Poisson Tail

- Large Deviation Regime: When $t \gg \mu$, we are in the so called "Large Deviation Regime", in which case the following holds

$$t^{-t} = e^{-t\ln t} \gg e^{-t^2/2}$$



- Hence, for $t \gg \mu$, the tail $\Pr\{X \geq t\}$ decays with rate at most $e^{-t\ln t}$, which is worse than the Gaussian tail, which decays with $e^{-t^2/2}$.



Small Deviation Regime of the Poisson Tail

- Let t be very close to μ. We can write this condition as

$$t = (1+\delta)\mu, \text{ for } 0 \leq \delta \leq 1$$

- This domain is called the “Small Deviation Regime”.
- In this regime, the tail simplifies to

$$\begin{aligned}
&\Pr\{S_n \geq (1+\delta)\mu\} \leq e^{-\mu}\left(\frac{e\mu}{(1+\delta)\mu}\right)^{(1+\delta)\mu} = e^{-\mu}\left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} \\
&= e^{-\mu}\left(\frac{e}{1+\delta}\right)^{\mu}\left(\frac{e}{1+\delta}\right)^{\delta\mu} = (1+\delta)^{-(1+\delta)\mu}e^{\delta\mu} \\
&= \exp\Big(\delta\mu - \mu(1+\delta)\ln(1+\delta)\Big) = \exp\Big(\mu\big(\delta - (1+\delta)\ln(1+\delta)\big)\Big)
\end{aligned} \tag{6}$$

- Note that $\delta - (1+\delta)\ln(1+\delta) < 0$ for $0 \leq \delta \leq 1$. Hence, we need an upper bound on $\delta - (1+\delta)\ln(1+\delta)$ in order to upper bound (6). We do this using Taylor series in the following.

Small Deviation Regime of the Poisson Tail

- From Taylor series expansion of $h(\delta)$, we know that $\exists c$, $0 \le c \le \delta \le 1$, such that for $h(\delta) = \delta - (1+\delta)\ln(1+\delta)$, we have

$$h(\delta) = h(0) + \delta h^{'}(0) + \frac{\delta^2}{2} h^{''}(c) \le -\frac{\delta^2}{2}$$

- Here $h(0) = 0$, $h^{'}(0) = 0$, $h^{''}(c) = -\frac{1}{1+c} \le -1/2$ for $0 \le c \le \delta \le 1$
- As a result,

$$\exp\Big(\mu\big(\delta - (1+\delta)\ln(1+\delta)\big)\Big) \le e^{-\mu\frac{\delta^2}{4}}$$

leading to the final result

$$\Pr\{S_n \ge (1+\delta)\mu\} \le e^{-\mu\frac{\delta^2}{4}}, \text{ for } 0 \le \delta \le 1 \tag{7}$$

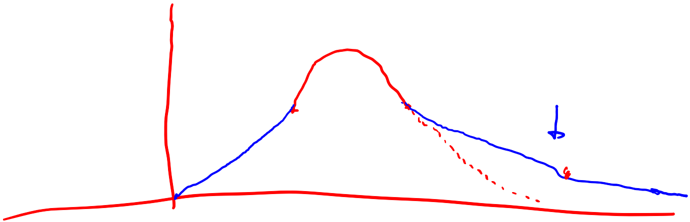
or written equivalently to

$$\Pr\{S_n - \mu \ge \delta\mu\} \le e^{-\mu\frac{\delta^2}{4}}, \text{ for } 0 \le \delta \le 1 \tag{8}$$

Small Deviation Regime of the Poisson Tail

- Using (8), prove at home that

$$\Pr\{|S_n - \mu| \geq \delta\mu\} \leq 2e^{-\mu\frac{\delta^2}{4}}, \text{ for } 0 \leq \delta \leq 1 \tag{9}$$



SDR vs LDR of the Poisson Tail

- When we compare the Small Deviation Regime (SDR) of the Poisson tail, we see that it behaves similar as Gaussian since it decays with rate at most

$$\Pr\{U \geq (1+\delta)\mu\} \leq e^{-\mu\frac{\delta^2}{4}}, \text{ for } 0 \leq \delta \leq 1 \tag{10}$$

- When we compare the Large Deviation Regime (LDR) of the Poisson tail, we see that it behaves worse than the Gaussian since it decays with rate at most

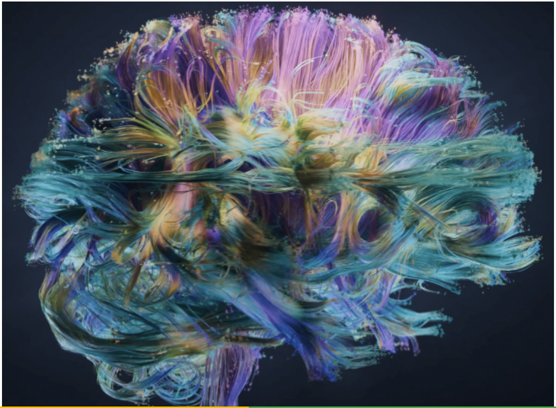
$$\Pr\{U \geq t\} \leq c_1 e^{-c_2 t \ln(t)}, \text{ for } t \gg \mu, \tag{11}$$

for some constants c_1 and c_2.

- Until now we only derived the one sided tail, i.e., $\Pr\{X \geq t\}$, for $t \geq \mu$. Derive at home an upper bound on the tail from the other side, i.e., $\Pr\{X \leq t\}$ for $t \leq \mu$. Note in your derivations that the Poisson distribution is not a symmetric distribution.

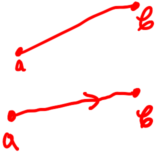
Application: Analysis of Complex Networks

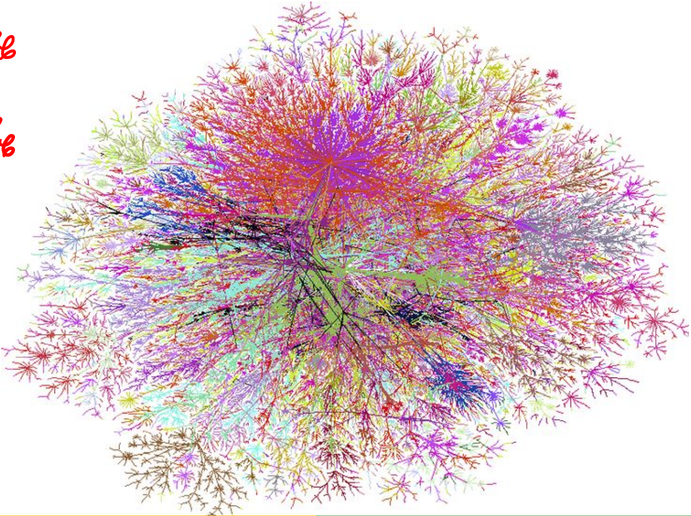
- Network is a graph. A graph consists of vertices (nodes) and edges/links (connections between nodes)
- Networks can be particularity complex,
- Example of neural connections in the brain as a network, where neurons are vertices and axons are edges:



Application: Analysis of Complex Networks

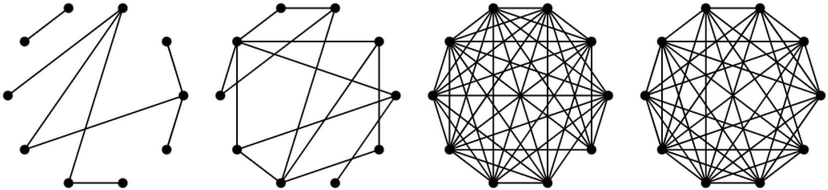
- Example of the internet as a network, where pages are vertices and links are edges





A Mathematical Model of Social Networks

- Social Network: Each person is a node. A link between two nodes exists if two people know each other.
- Let's try to generate the simplest form of a Social Network that we can work with mathematically and try to understand its properties.
- Def (Random Graph): Fix n nodes. Connect each pair of nodes independently with probability p. Then, this network is known as the random network, denoted by $G(n, p)$.
- Example of random graphs with 10 nodes and increasing p from left to right:



A Mathematical Model of Social Networks

- The "degree", denoted by $\deg(i)$, is the number of connected edges of node i, i.e, the number of connections of a node i
- What is the mean of $\deg(i)$, also known as the "average degree"? This can be derived from how the random graph is created: a) each nodes makes connections independently, b) there are $n-1$ nodes in total with which node i can make a connection. Hence,

$$\deg(i) = \sum_{i=1}^{n-1} X_i,$$

$p = \frac{d}{n-1}$

where X_i are i.i.d. and $X_i \sim \mathrm{Ber}(p)$, $\forall i$
- Then, $E[\deg(i)] = (n-1)p = d$, is the average degree of each node.
- We will prove the following phenomenon, where c's are constants:
  - If $d < c_1 \ln(c_2 n)$, then there are few nodes with very large number of connections and most of the other nodes have only few connections and are thus very isolated.
  - If $d > c_3 \ln(c_4 n)$, then all of the nodes have almost the same number of connections.

A Mathematical Model of Social Networks

- Def (Regularity): A graph is called d-regular if $\deg(i) = d, \forall i$.
- We will show that for random graphs, by increasing d, we obtain a phase transition from non-regular to approximately regular graph that occurs around $d \approx \ln(n)$.
- Specifically, if $d \overset{c_1}{\gg} \ln(\overset{c_2}{n})$, then the random graph is almost d-regular. Hence, all nodes have almost the same number of connections.
- If $d \overset{c_3}{\ll} \ln(\overset{c_4}{n})$, then the random graph is very far from regular. In this case, there are few nodes with very high number of connections, which form hubs, and the majority of the nodes are very isolated with only few connections.
- There is also a big part when both $d \overset{c_1}{\gg} \ln(\overset{c_2}{n})$ and $d \overset{c_3}{\ll} \ln(\overset{c_4}{n})$ do not hold, which we will not look at, but is also an important region of d for investigating social networks.

A Mathematical Model of Social Networks

- First recall that $d = E[\deg(i)] = (n-1)p$.
- Let's fix d. Then, $p = d/(n-1)$.
- Thm (Regularity): A random graph $G(n,p)$ with

$$p = \frac{d}{n-1}$$

is d-regular with probability $1-\epsilon$ if d satisfies

$$d \geq \frac{4}{\delta^2} \ln\left(\frac{2n}{\epsilon}\right). \tag{12}$$

More precisely, if (12) holds, then the following holds for each node i

$$\Pr\{|\deg(i) - d| \leq \delta d\} \geq 1 - \epsilon, \forall i, \tag{13}$$

where δ satisfies $0 \leq \delta \leq 1$.

A Mathematical Model of Social Networks

- Proof: Since (13) needs to hold $\forall i$, i.e., for all n nodes, we need to prove

$$\Pr\left\{\bigcap_{i=1}^{n}(|\deg(i)-d|)\leq\delta d\right\}\leq 1-\epsilon \tag{14}$$

On the other hand

$$\Pr\left\{\bigcap_{i=1}^{n}(|\deg(i)-d|)\leq\delta d\right\}=1-\Pr\left\{\bigcup_{i=1}^{n}(|\deg(i)-d|)\geq\delta d\right\},$$

AND

OR

which comes from de Morgan's law, given by

$$\Pr\left\{\bigcap_{i=1}^{n}A_i\right\}=1-\Pr\left\{\bigcup_{i=1}^{n}A_i^c\right\}$$

that holds for any events A_i and their compliments A_i^c.
Hence, for (14) to hold, we need to prove that the following holds

$$\Pr\left\{\bigcup_{i=1}^{n}(|\deg(i)-d|)\geq\delta d\right\}\leq\epsilon \tag{15}$$

A Mathematical Model of Social Networks

$$\Pr\left\{\bigcup_{i=1}^{n}(|\deg(i)-d|)\geq\delta d\right\}\overset{(a)}{\leq}\sum_{i=1}^{n}\Pr\{(|\deg(i)-d|)\geq\delta d\}\overset{(b)}{\leq}\sum_{i=1}^{n}2e^{-d\frac{\delta^2}{4}}$$

$$=n2e^{-d\frac{\delta^2}{4}}=e^{-d\frac{\delta^2}{4}+\ln(n)+\ln(2)},$$

$e^{\ln(n)}$ $e^{\ln(2)}$

where (a) is due to the union bound which states for any events A_i's

$$\boxed{\Pr\left\{\bigcup_{i=1}^{n}A_i\right\}\leq\sum_{i=1}^{n}\Pr\{A_i\}}$$

and (b) is due to (9).

Now, to prove (15), we need

$$e^{-d\frac{\delta^2}{4}+\ln(n)+\ln(2)} \leq \epsilon, \qquad \Big|\, -\ln(\epsilon)$$

which occurs if

$$d\frac{\delta^2}{4} - \ln(n) - \ln(2) \geq -\ln(\epsilon),$$

or equivalently if

$$d \geq \frac{4}{\delta^2} \ln\left(\frac{2n}{\epsilon}\right), \qquad \text{Q.E.D.}$$

A Mathematical Model of Social Networks

- Thm (Sparsity): A random graph $G(n, p)$ with

$$p = \frac{d}{n-1} \tag{16}$$

sparse

is far from d-regular if

$$\frac{2}{n} \ln\left(\frac{1}{\epsilon}\right) \frac{1}{1 - \frac{1}{\gamma} - \frac{1}{\gamma}\ln(\gamma)} \leq d \leq 2 \frac{1}{1 + \gamma \ln(\gamma)} \ln\left(n \frac{1}{\ln(1/\epsilon)}\right) \tag{17}$$

More precisely, for a random graph $G(n, p)$ with p given by (16) satisfies

$$\Pr\{\exists i : \deg(i) \geq \gamma d\} \geq 1 - \epsilon, \tag{18}$$

and

$$\Pr\left\{\exists i : \deg(i) \leq \frac{d}{\gamma}\right\} \geq 1 - \epsilon, \tag{19}$$

where γ satisfies $\gamma \gg 1$, if d satisfies (17).

A Mathematical Model of Social Networks

- Proof: To prove (18), we need to prove that the following holds for the given conditions

10 → 100 1

$$\Pr\left\{\bigcup_{i=1}^{n} \deg(i) \geq \gamma d\right\} \geq 1-\epsilon \tag{20}$$

To prove (19), we need to prove that the following holds for the given conditions

$$\Pr\left\{\bigcup_{i=1}^{n} \deg(i) \leq \frac{d}{\gamma}\right\} \geq 1-\epsilon \tag{21}$$

A Mathematical Model of Social Networks

- To prove (18), we will use the reverse Chernoff bound of $S_n = \sum_{i=1}^{n-1} X_i$, where $X_i \sim \text{Ber}(d/n)$ are i.i.d., given by

$$\Pr\{S_n \geq t\} \geq e^{-\mu}\left(\frac{\mu}{t}\right)^t \qquad (22)$$

3 pts

Prove (22) **at home. The first student that sends me the proof of** (22)**, I will give that student points! Then, for each subsequent proof that a student sends me, I will give points only if the proof of** (22) **is different than the proofs sent by the previous students.**

- Substituting S_n with $\deg(i)$, μ with d, and t with γd in (22), we obtain

$$\Pr\{\deg(i) \geq \gamma d\} \geq e^{-d}\left(\frac{d}{\gamma d}\right)^{\gamma d} = e^{-d}\gamma^{-\gamma d} = e^{-d}e^{-\gamma d \ln(\gamma)} = e^{-d(1+\gamma\ln(\gamma))} \qquad (23)$$

We can write the left-hand side of (20), as

$$\Pr\left\{\bigcup_{i=1}^{n} \deg(i) \geq \gamma d\right\} = 1 - \Pr\left\{\bigcap_{i=1}^{n} \deg(i) \leq \gamma d\right\} \geq 1 - \epsilon \quad (24)$$

Hence, to prove (20), we need to prove that

$$\Pr\left\{\bigcap_{i=1}^{n} \deg(i) \leq \gamma d\right\} \leq \epsilon \quad (25)$$

Now we can write (25) as

$$
\begin{aligned}
\Pr\left\{\bigcap_{i=1}^{n} \deg(i) \leq \gamma d\right\} &\overset{(a)}{=} \prod_{i=1}^{n} \Pr\left\{\deg(i) \leq \gamma d\right\} \\
&= \prod_{i=1}^{n} (1 - \Pr\left\{\deg(i) \geq \gamma d\right\}) \\
&\overset{(b)}{\leq} \prod_{i=1}^{n} \left(1 - e^{-d(1+\gamma \ln(\gamma))}\right) \\
&= \left(1 - e^{-d(1+\gamma \ln(\gamma))}\right)^{n}, \qquad (26)
\end{aligned}
$$

where (a) comes from the independence of the degrees at each nodes and (b) comes from (23).

A Mathematical Model of Social Networks

Comparing (25) and (26), we see that we need to have

- We want

$$\left(1 - e^{-d(1+\gamma \ln(\gamma))}\right)^n \leq \epsilon,$$

which is equivalent to

$$1 - e^{-d(1+\gamma \ln(\gamma))} \leq \epsilon^{1/n}$$

which is equivalent to

$$-d(1 + \gamma \ln(\gamma)) \leq \ln(1 - \epsilon^{1/n})$$

and finally to

$$d \leq -\frac{1}{1 + \gamma \ln(\gamma)} \ln(1 - \overbrace{\epsilon^{1/n}}^{<1}) \tag{27}$$

A Mathematical Model of Social Networks

We want to simplify (27) even further. To this end, we use

$$-\ln\left(1-\epsilon^{1/n}\right) = -\ln\left(1-e^{\frac{\ln(\epsilon)}{n}}\right)$$

Now using the Taylor Series expansion $e^x = \sum_{k=0}^{\infty} x^k/k!$, we obtain the bound

$$e^{\frac{\ln(\epsilon)}{n}} \geq 1 + \frac{\ln(\epsilon)}{n}$$

Inserting this bound above, we obtain

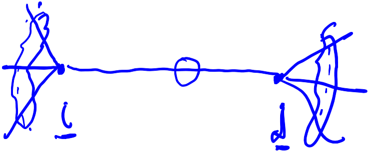
$\ln(\frac{1}{\epsilon})/$

$$-\ln\left(1-\epsilon^{1/n}\right) \geq -\ln\left(1-1-\frac{\ln(\epsilon)}{n}\right) = \ln\left(\frac{n}{\ln(1/\epsilon)}\right) \qquad (28)$$

Inserting (37) into (27), we obtain (17)
Q.E.D.

A Mathematical Model of Social Networks

- Did we made a mistake in (26) in (a) by assuming independence?



$d_{old} < th/2$

$d_{ew} < th$

$d_{new} < \frac{th}{2}$

$n/2$

d

$\boxed{d_{new} = \frac{d}{2}}$

A Mathematical Model of Social Networks

- Now let's prove (21)

$$\Pr\left\{\bigcup_{i=1}^{n}\deg(i)\le\frac{d}{\gamma}\right\}=1-\Pr\left\{\bigcap_{i=1}^{n}\deg(i)\ge\frac{d}{\gamma}\right\}\ge 1-\epsilon \tag{29}$$

Hence, we need to prove that

$$\Pr\left\{\bigcap_{i=1}^{n}\deg(i)\ge\frac{d}{\gamma}\right\}\le\epsilon \tag{30}$$

$$\Pr\left\{\bigcap_{i=1}^{n}\deg(i)\ge\frac{d}{\gamma}\right\}\overset{(a)}{=}\prod_{i=1}^{n}\Pr\left\{\deg(i)\ge\frac{d}{\gamma}\right\}\overset{(a)}{\le}\prod_{i=1}^{n}e^{-d}\left(\frac{ed}{d/\gamma}\right)^{d/\gamma}$$
$$=e^{-nd}e^{nd/\gamma}\gamma^{nd/\gamma}=e^{-nd(1-1/\gamma)}e^{nd/\gamma\ln(\gamma)}=e^{-nd\left(1-\frac{1}{\gamma}-\frac{1}{\gamma}\ln(\gamma)\right)} \tag{31}$$

where (a) comes from (5) by replacing μ with d and t with d/γ.

A Mathematical Model of Social Networks

Hence, we need

$$e^{-nd\left(1-\frac{1}{\gamma}-\frac{1}{\gamma}\ln(\gamma)\right)} \leq \epsilon$$

Or equivalently

$$nd\left(1-\frac{1}{\gamma}-\frac{1}{\gamma}\ln(\gamma)\right) \geq \ln\left(\frac{1}{\epsilon}\right)$$

Finally, we get

$$d > \frac{2}{n}\ln\left(\frac{1}{\epsilon}\right)\frac{1}{1-\frac{1}{\gamma}-\frac{1}{\gamma}\ln(\gamma)} \qquad (32)$$

where the right hands side converges fast to 0 as n grows. Hence, for medium to large n, the random graph will always have isolated nodes with very few connections. Q.E.D.

Note that 2 in is due to the independence trick.

A Mathematical Model of Social Networks

- Stop to think what have we proved with the last theorem:
  - 1. We proved that for small enough d, there are some nodes, but few in numbers, that have large number of connections
  - We proved that for small enough d, there are nodes, in fact a large number of them, that have only few number of connections and are thus very isolated.