A Mathematical Model for Finding the Culprit Who Spreads Rumors

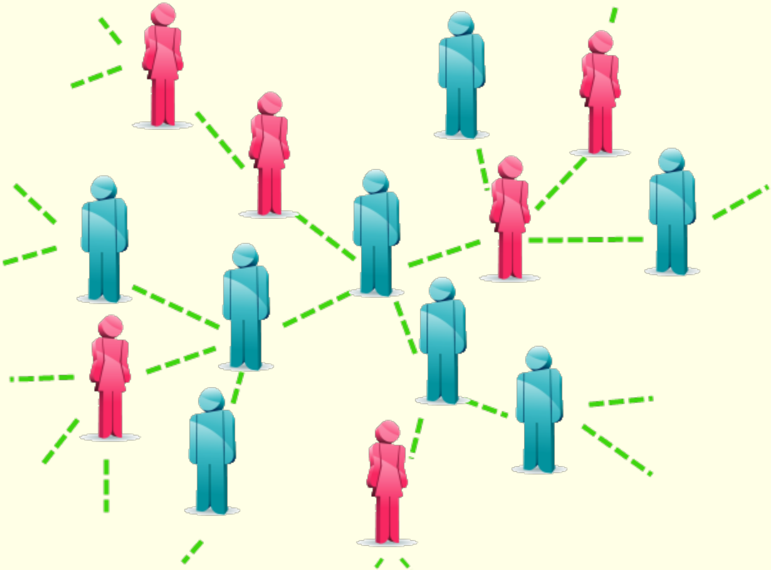
搜尋散佈謠言者的數學模型

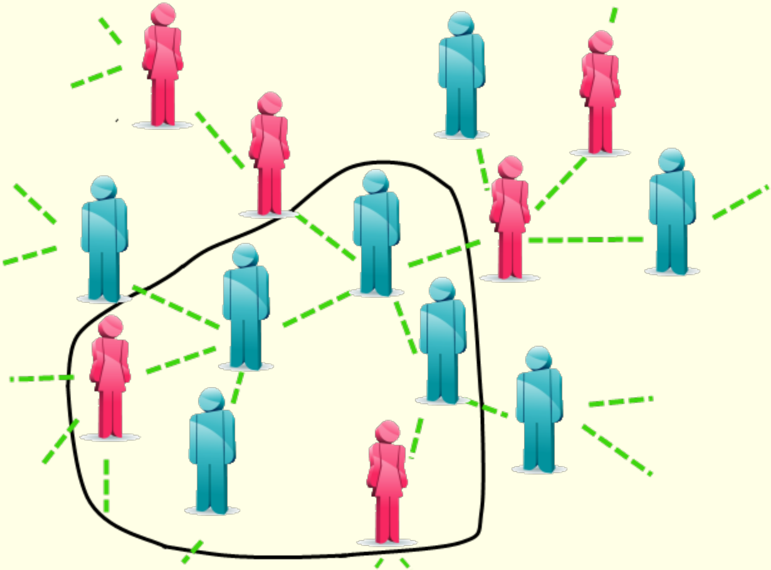
研究生: 李姿慧

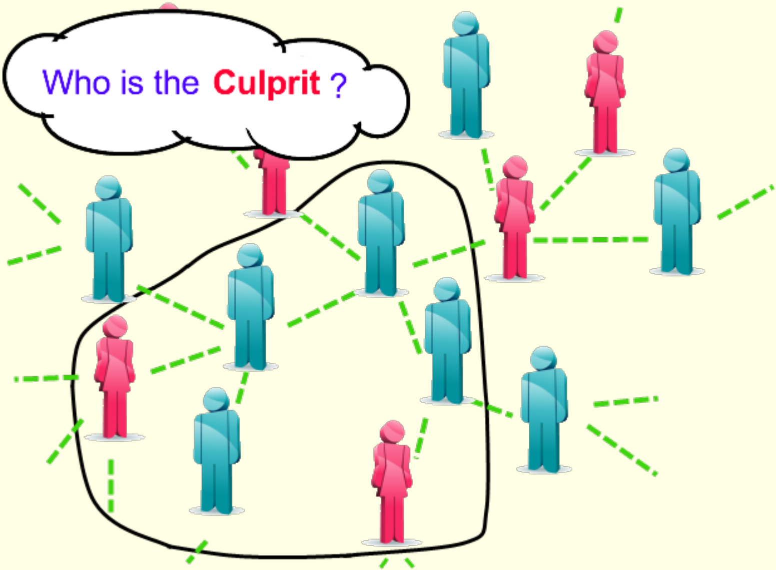
指導教授: 傅恆霖 教授

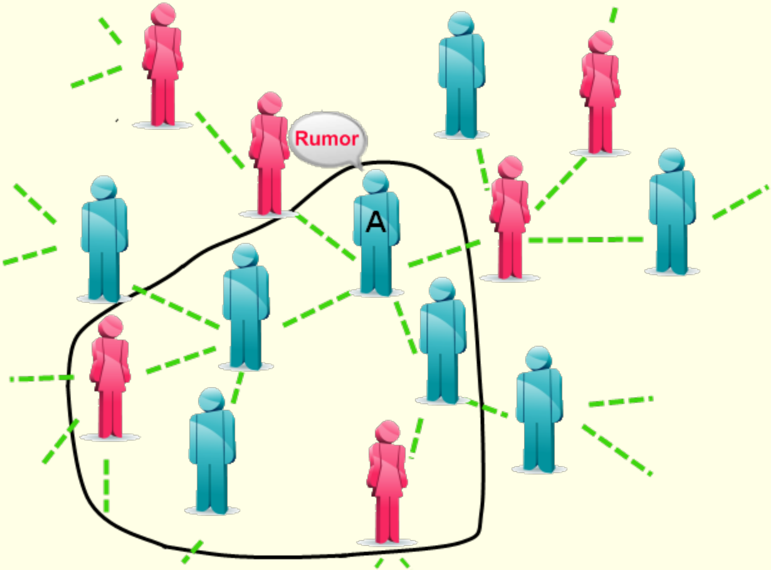
Department of Applied Mathematics
National Chiao Tung University

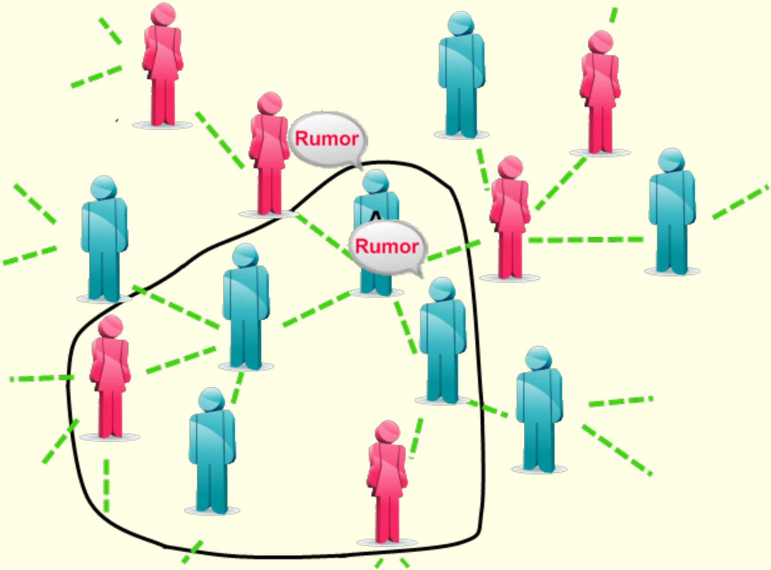
June 12, 2012

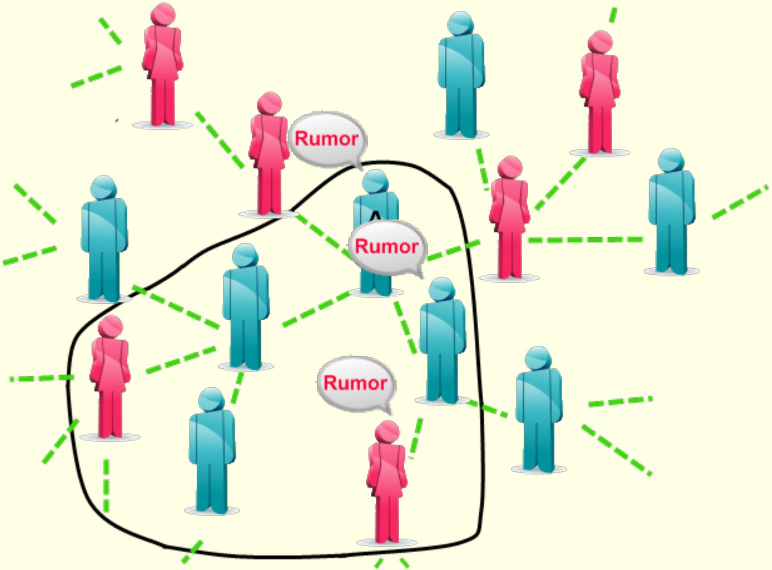


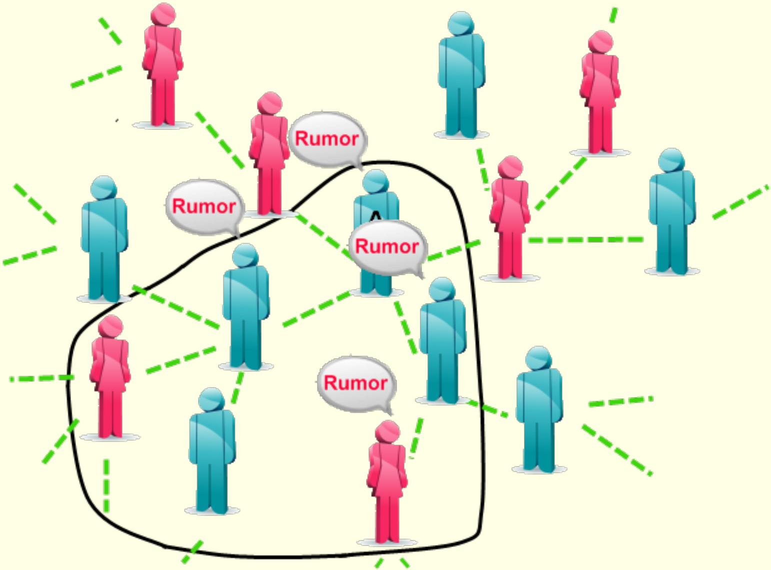


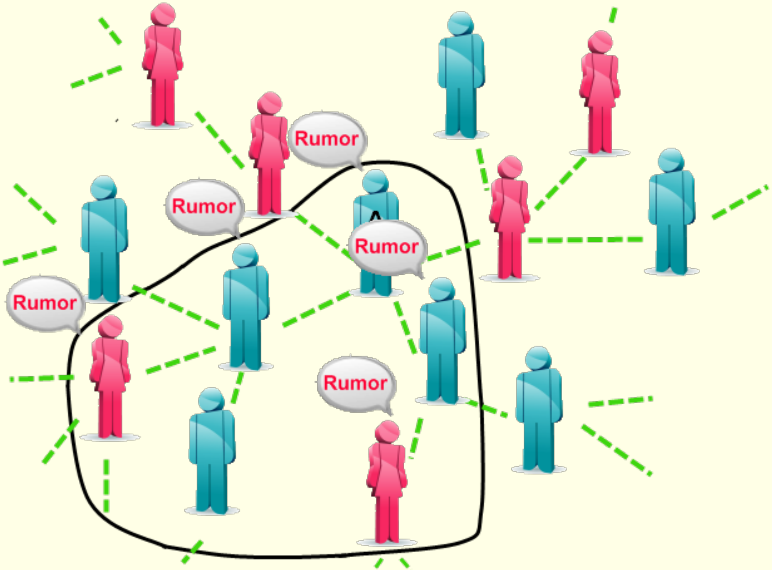

Who is the Culprit ?

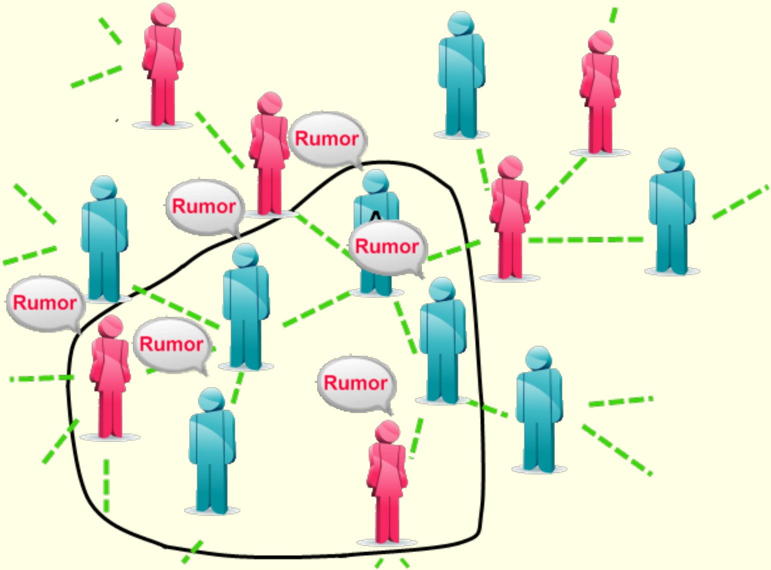

Rumor
A

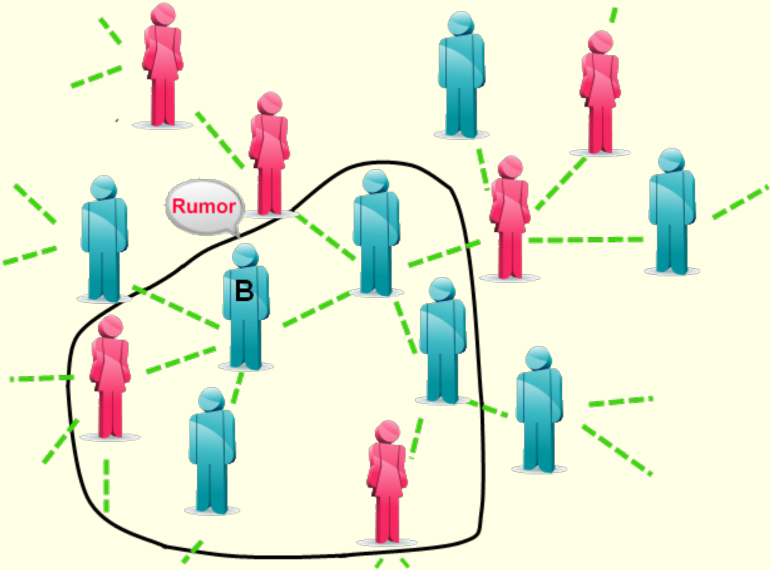

Rumor
Rumor

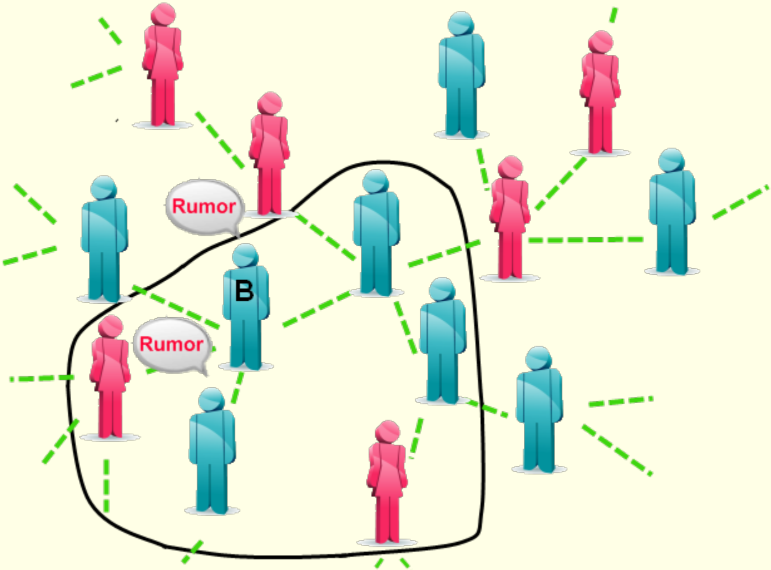

Rumor
Rumor
Rumor

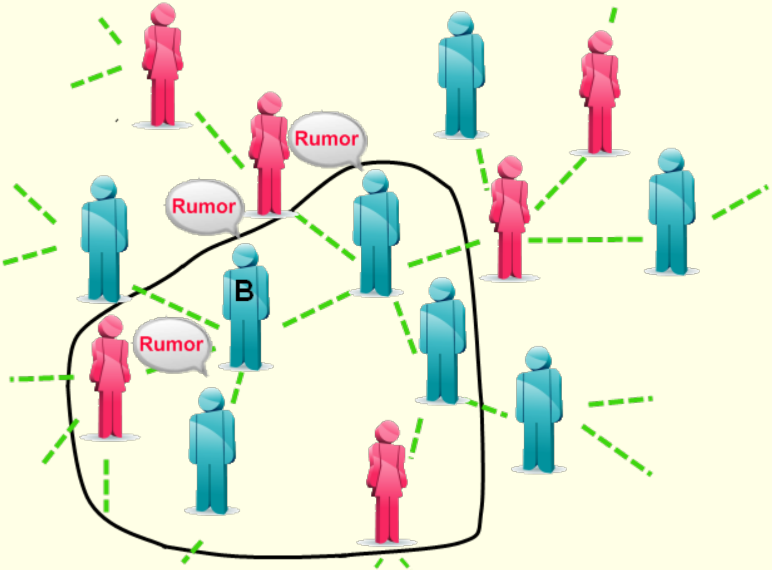

Rumor
Rumor
Rumor
Rumor

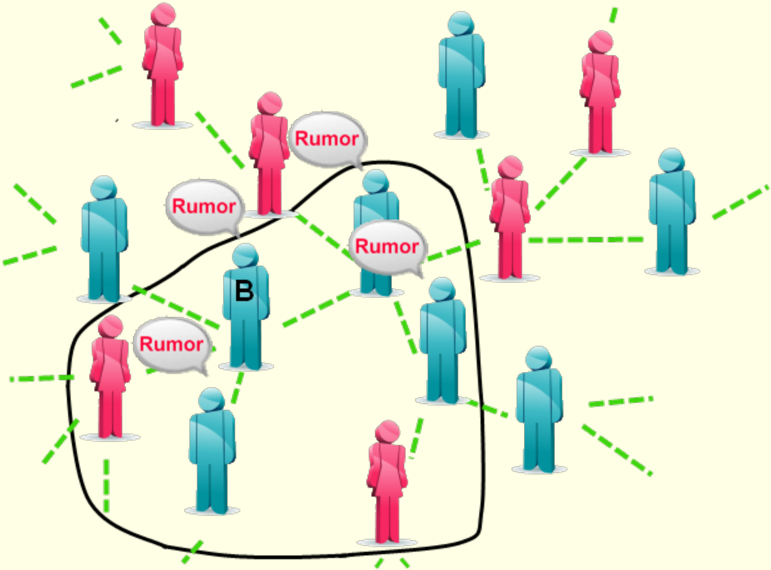

Rumor
Rumor
Rumor
Rumor
Rumor

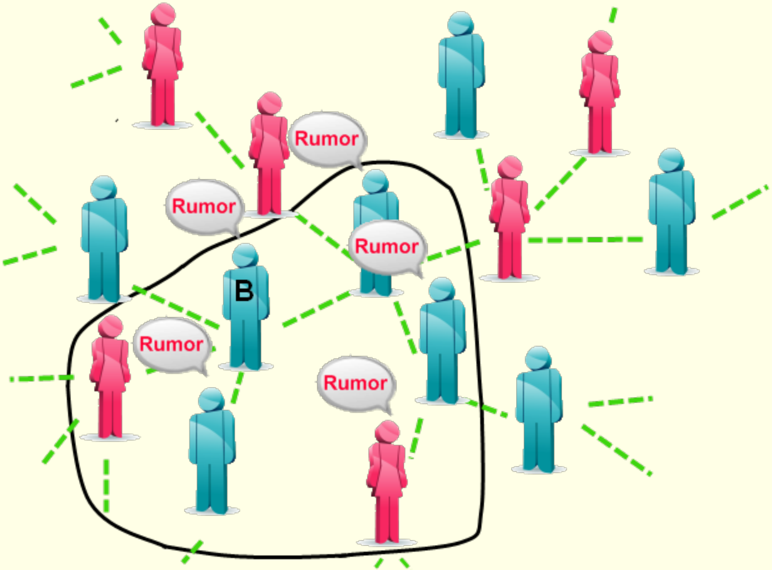

Rumor
Rumor
Rumor
Rumor
Rumor
Rumor

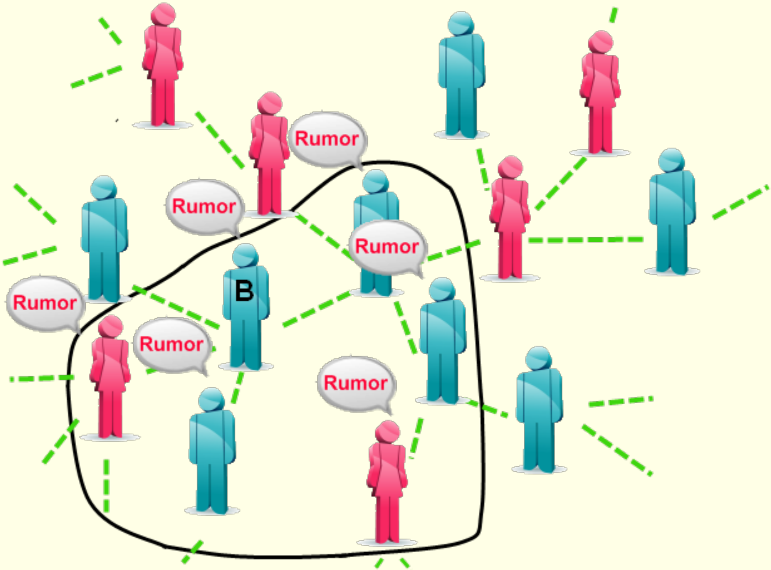

Rumor
B

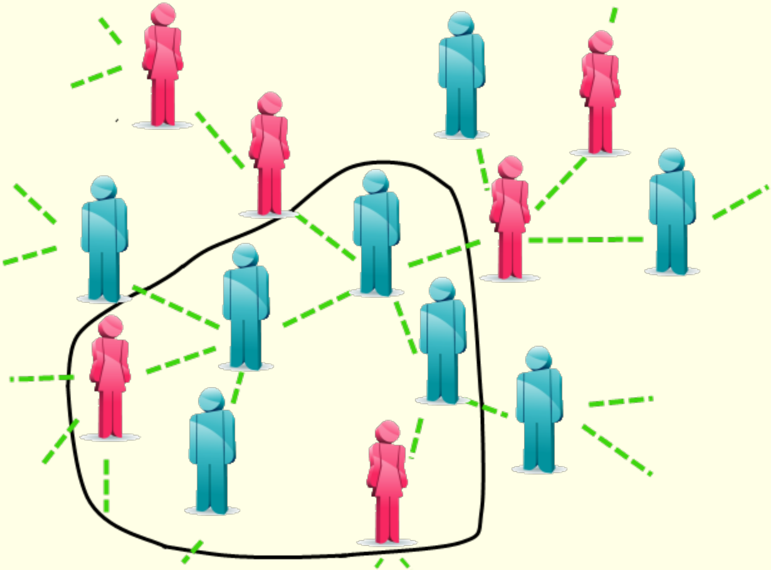

Rumor
B
Rumor

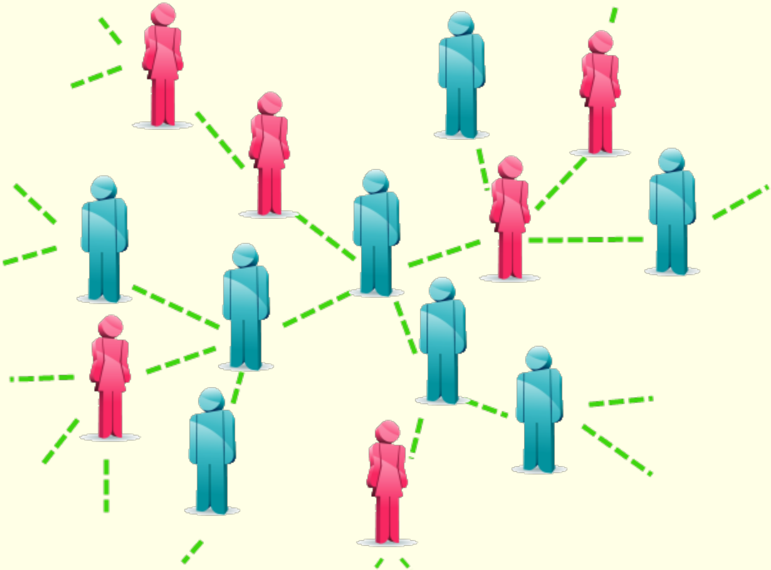

Rumor
Rumor
B
Rumor


Rumor
Rumor
Rumor
Rumor
B

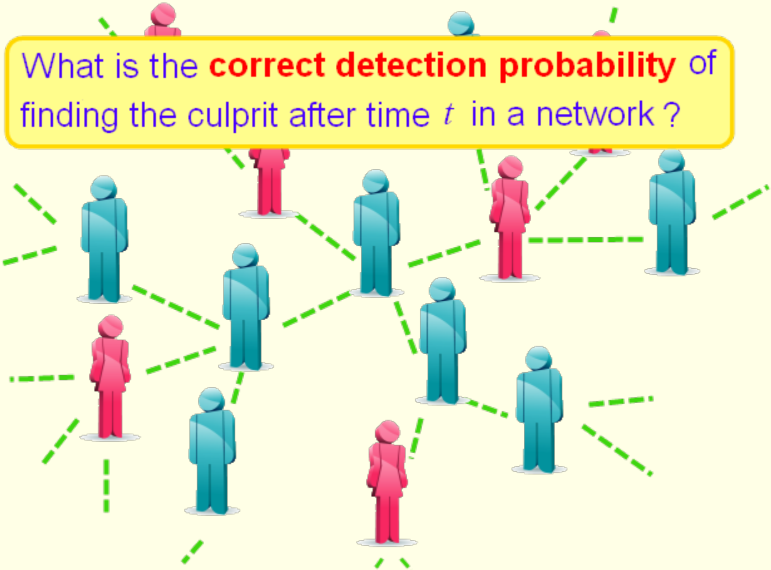

Rumor
Rumor
Rumor
B
Rumor
Rumor


Rumor
Rumor
Rumor
Rumor
Rumor
B
Rumor





What is the **correct detection probability** of finding the culprit after time t in a network ?



SI model (Kermack and McKendrick, 1927)

- An epidemiological model.

SI model (Kermack and McKendrick, 1927)

- An epidemiological model.
- In a closed population.

SI model (Kermack and McKendrick, 1927)

- An epidemiological model.
- In a closed population.
- Two types of people:
  - S_t: The number of susceptible (可被感染的) people at time t.
  - I_t: The number of infected (已被感染的) people at time t.

SI model (Kermack and McKendrick, 1927)

- An epidemiological model.
- In a closed population.
- Two types of people:
  - S_t: The number of susceptible (可被感染的) people at time t.
  - I_t: The number of infected (已被感染的) people at time t.
- $I_{t+\Delta t} - I_t = \beta S_t I_t \Delta t$.

SI model (Kermack and McKendrick, 1927)

- An epidemiological model.
- In a closed population.
- Two types of people:
  - S_t: The number of susceptible (可被感染的) people at time t.
  - I_t: The number of infected (已被感染的) people at time t.
- $I_{t+\Delta t} - I_t = \beta S_t I_t \Delta t$.
- A discrete-time model:

$$\begin{cases} S_{t+1} = S_t - \beta S_t I_t, \\ I_{t+1} = I_t + \beta S_t I_t, \\ I_0 = \text{A constant}. \end{cases}$$

Rumor spreading model

- S_t: The number of people who don't know rumors yet.
- I_t: The number of people who have known rumors.

Rumor spreading model

- S_t: The number of people who don't know rumors yet.
- I_t: The number of people who have known rumors.

Assumption

Using a fixed population, $S_t + I_t = N$.

Assume $S_0 = N$, $I_0 = 0$, and

$$S_{t+1} = S_t - 1, \quad I_{t+1} = I_t + 1,$$

Rumor spreading model

- S_t: The number of people who don't know rumors yet.
- I_t: The number of people who have known rumors.

Assumption

Using a fixed population, $S_t + I_t = N$.

Assume $S_0 = N$, $I_0 = 0$, and

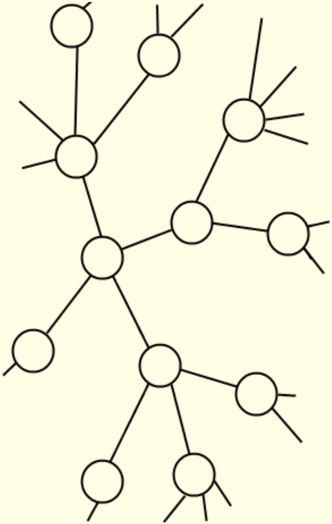
$$S_{t+1} = S_t - 1, \quad I_{t+1} = I_t + 1,$$

- $V(G)$: People.
- $E(G)$: Relationship between two people.

Rumor spreading model

Let G_t be a subgraph of order t of G.

This graph is compose of t infected vertices which are people who have known rumors at time t.

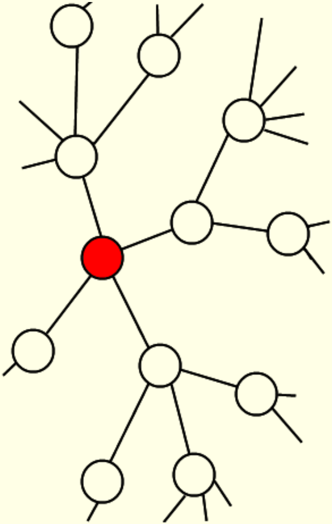


Rumor spreading model

Let G_t be a subgraph of order t of G.

This graph is compose of t infected vertices which are people who have known rumors at time t.

- G_1: **Rumor source**.



Rumor spreading model

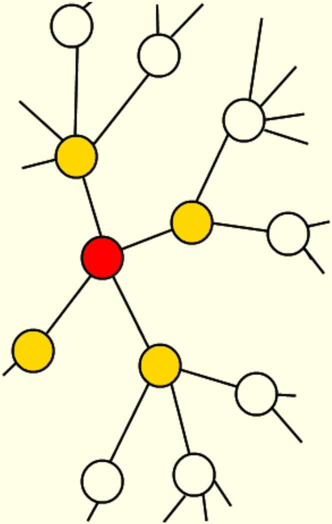
Let G_t be a subgraph of order t of G.

This graph is compose of t infected vertices which are people who have known rumors at time t.

- G_1: **Rumor source**.

In each discrete time-step $t+1$, $t>0$, G_{t+1} develops from G_t by adding a vertex z with an edge with the following probability distribution:

$$\mathbf{P}_{t+1}(z) = \frac{1}{\sum_{v \in V(G_t)} d(v) - 2(t-1)}. \qquad (1)$$

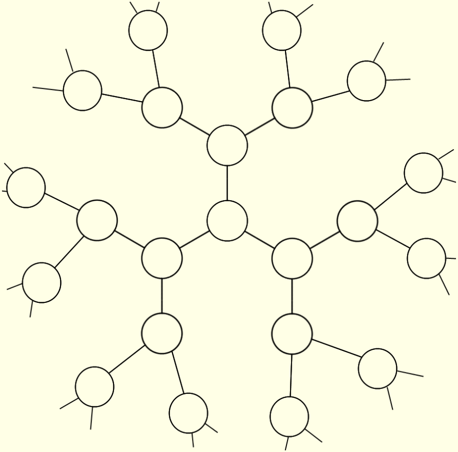


d-regular tree

- **Countably infinite** vertex set.
- **Every** vertex has d neighbors.

  $N(v)$: Neighborhood of v.

  $d(v)$: Degree of v.

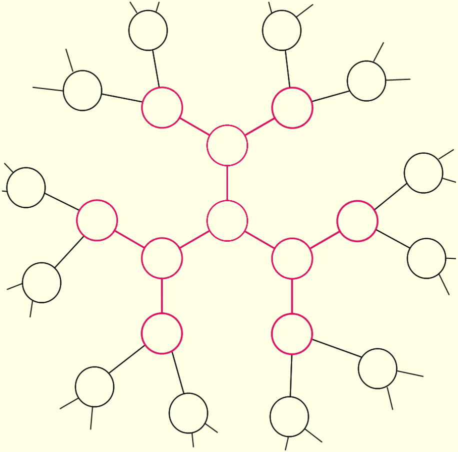


d-regular tree

- **Countably infinite** vertex set.
- **Every** vertex has d neighbors.

  $N(v)$: Neighborhood of v.

  $d(v)$: Degree of v.

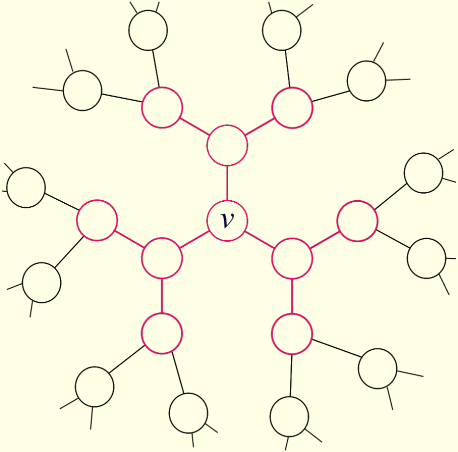


d-regular tree

- **Countably infinite** vertex set.
- **Every** vertex has d neighbors.

  $N(v)$: Neighborhood of v.

  $d(v)$: Degree of v.
- A rooted tree T_v can be **decomposed** into d subtrees.



d-regular tree

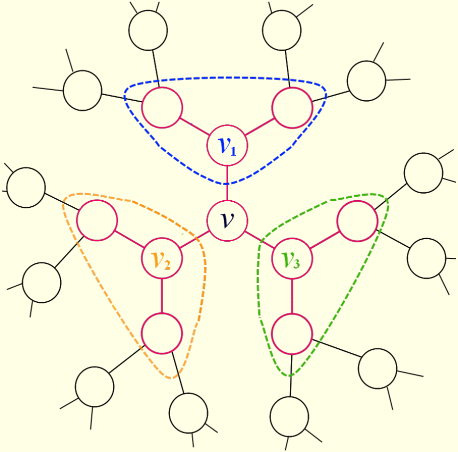
- **Countably infinite** vertex set.
- **Every** vertex has d neighbors.

  $N(v)$: Neighborhood of v.

  $d(v)$: Degree of v.

- A rooted tree T_v can be **decomposed** into d subtrees.

$(T^v_{v_1}, T^v_{v_2}, T^v_{v_3})$: Branches of T_v, and $t^v_{v_1} + t^v_{v_2} + t^v_{v_3} = t_v - 1$.



Definition 1

Consider **rumor centrality**(謠言向心性) $R(v, G_n)$, the followings are equivalent.

Definition 1

Consider **rumor centrality**(謠言向心性) $R(v, G_n)$, the followings are equivalent.

- The number of distinct ways to spread a rumor to every vertex in G_n with v as the source.

Definition 1

Consider **rumor centrality**(謠言向心性) $R(v, G_n)$, the followings are equivalent.

- The number of distinct ways to spread a rumor to every vertex in G_n with v as the source.
- **Rumor center** of G_n is the vertex with maximum rumor centrality.

Definition 1

Consider **rumor centrality**(謠言向心性) $R(v, G_n)$, the followings are equivalent.

- The number of distinct ways to spread a rumor to every vertex in G_n with v as the source.
- **Rumor center** of G_n is the vertex with maximum rumor centrality.

Proposition 1 (D. Shah and T. Zaman, 2011 [6])

Rumor center is the **maximum likelihood estimator**(最大概似估計) of a regular tree.

Definition 1

Consider **rumor centrality**(謠言向心性) $R(v, G_n)$, the followings are equivalent.

- The number of distinct ways to spread a rumor to every vertex in G_n with v as the source.
- **Rumor center** of G_n is the vertex with maximum rumor centrality.

Proposition 1 (D. Shah and T. Zaman, 2011 [6])

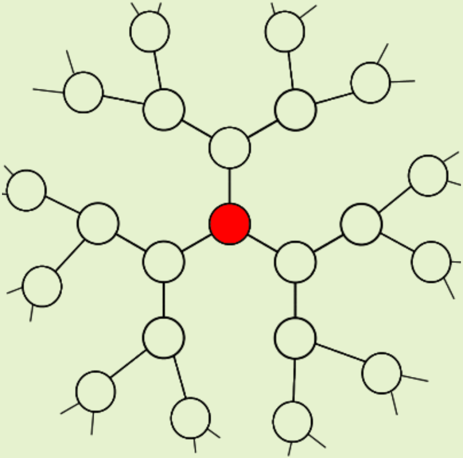
Rumor center is the **maximum likelihood estimator**(最大概似估計) of a regular tree.

Proposition 2 (D. Shah and T. Zaman, 2011 [6])

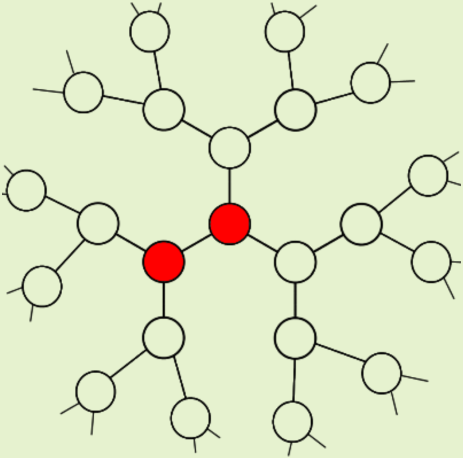
Given an n vertices tree,

- Node v is the **only** rumor center if and only if $t_u^v < \frac{n}{2}$ for all $u \neq v$.
- If there is a node v such that $t_u^v = \frac{n}{2}$, then u and v both are rumor centers.

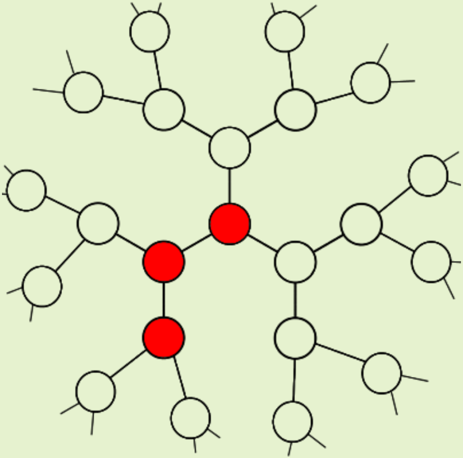
$d = 3, t = n = 5$



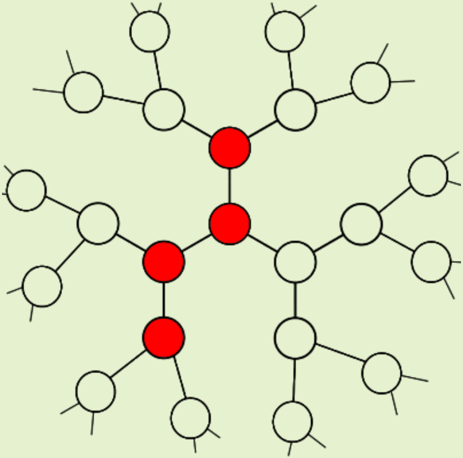
$d = 3, t = n = 5$



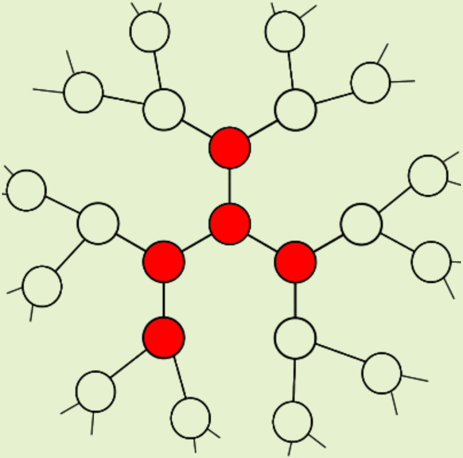
$d = 3, t = n = 5$



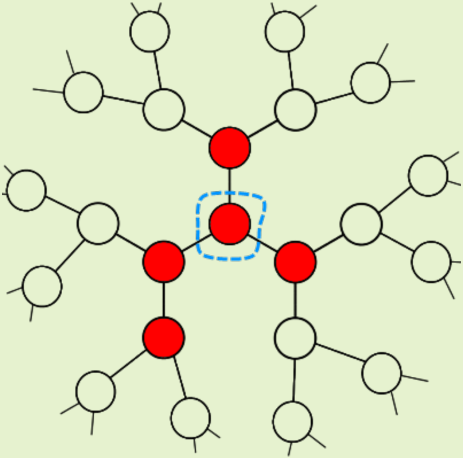
$d = 3, t = n = 5$



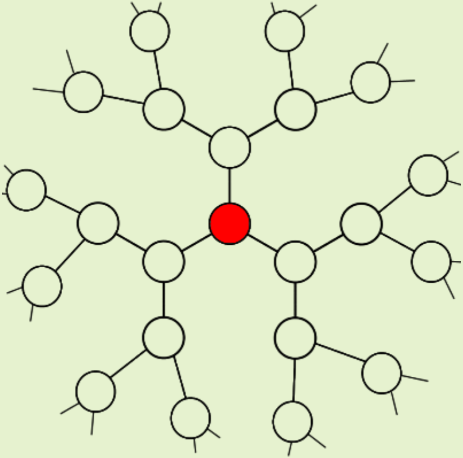
$d = 3, t = n = 5$



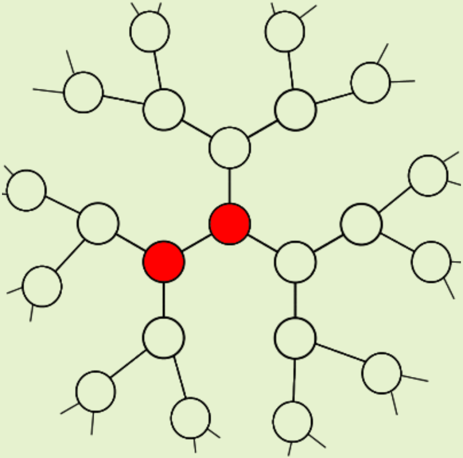
$d = 3, t = n = 5$



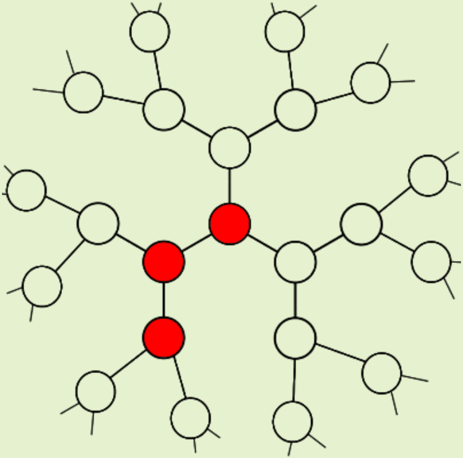
$d = 3, t = n = 5$



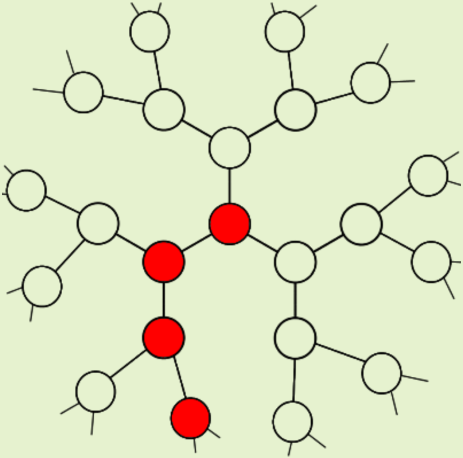
$d = 3, t = n = 5$



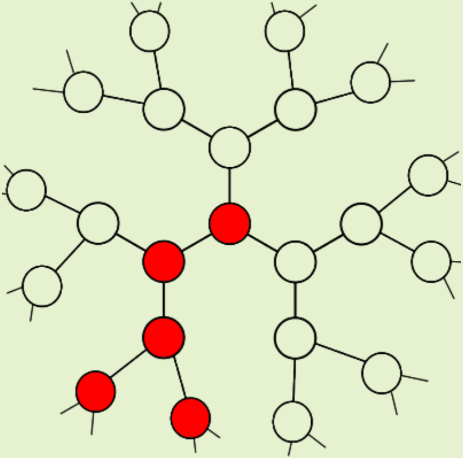
$d = 3, t = n = 5$



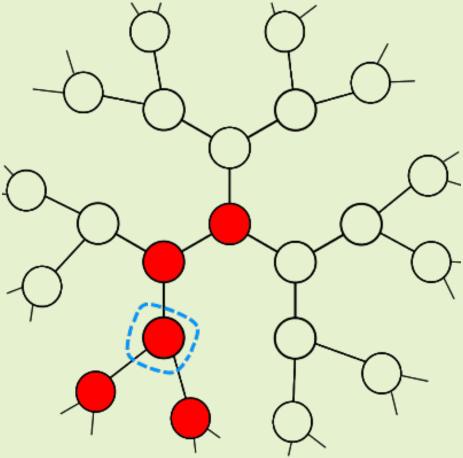
$d = 3, t = n = 5$



$d = 3, t = n = 5$



$d = 3, t = n = 5$



Definition 2

Let $E_t(G)$ be the event of correct rumor source detection under the ML rumor source estimator after time t on a graph G. If the graph G considered is prescribed , then we use E_t to denote $E_t(G)$.

Definition 2

Let $E_t(G)$ be the event of correct rumor source detection under the ML rumor source estimator after time t on a graph G. If the graph G considered is prescribed , then we use E_t to denote $E_t(G)$.

Theorem 1

Suppose the rumor has spread in a **2-regular tree**. Then we have that

$$\mathbf{P}(E_t) = O\left(\frac{1}{\sqrt{t}}\right).$$

Definition 2

Let $E_t(G)$ be the event of correct rumor source detection under the ML rumor source estimator after time t on a graph G. If the graph G considered is prescribed , then we use E_t to denote $E_t(G)$.

Theorem 1

Suppose the rumor has spread in a **2-regular tree**. Then we have that

$$\mathbf{P}(E_t) = O\left(\frac{1}{\sqrt{t}}\right).$$

Theorem 2

Suppose the rumor has spread in a **regular tree**. Then we have that

$$0 < \mathbf{P}\left(E_t\right) \leq \frac{1}{2}.$$

Definition 3

$$A_d = \{(a_1, a_2, \cdots, a_d) \mid 1 \leq a_i < \frac{n}{2}, \sum_{i=1}^{d} a_i = n - 1\}, \text{ and}$$

$$B_d = \{(b_1, b_2, \cdots, b_d) \mid b_i \in \mathbb{N}, \sum_{i=1}^{d} b_i = n - 1\}.$$

Clearly, $A_d \subseteq B_d$.

Definition 3

$$A_d = \{(a_1, a_2, \cdots, a_d) \mid 1 \le a_i < \frac{n}{2}, \sum_{i=1}^{d} a_i = n-1\}, \text{ and}$$

$$B_d = \{(b_1, b_2, \cdots, b_d) \mid b_i \in \mathbb{N}, \sum_{i=1}^{d} b_i = n-1\}.$$

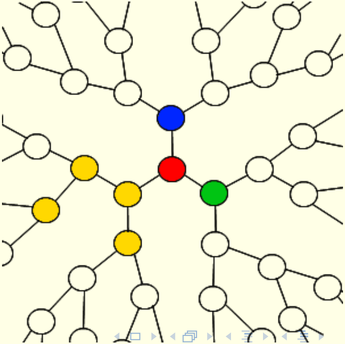
Clearly, $A_d \subseteq B_d$.

$d = 3, t = n = 7$

(1, 1, 4), (1, 4, 1), (4, 1, 1),

(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1),

(3, 1, 2), (3, 2, 1), (2, 2, 2).



Definition 3

$$A_d = \{(a_1, a_2, \cdots, a_d) | 1 \leq a_i < \frac{n}{2}, \sum_{i=1}^{d} a_i = n - 1\}, \text{ and}$$

$$B_d = \{(b_1, b_2, \cdots, b_d) | b_i \in \mathbb{N}, \sum_{i=1}^{d} b_i = n - 1\}.$$

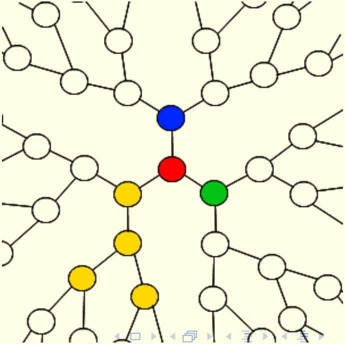
Clearly, $A_d \subseteq B_d$.

$d = 3, t = n = 7$

$(1,1,4), (1,4,1), (4,1,1),$

$(1,2,3), (1,3,2), (2,1,3), (2,3,1),$

$(3,1,2), (3,2,1), (2,2,2).$



Definition 3

$$A_d = \{(a_1, a_2, \cdots, a_d) | \, 1 \le a_i < \frac{n}{2}, \sum_{i=1}^{d} a_i = n - 1\}, \text{ and}$$

$$B_d = \{(b_1, b_2, \cdots, b_d) | \, b_i \in \mathbb{N}, \sum_{i=1}^{d} b_i = n - 1\}.$$

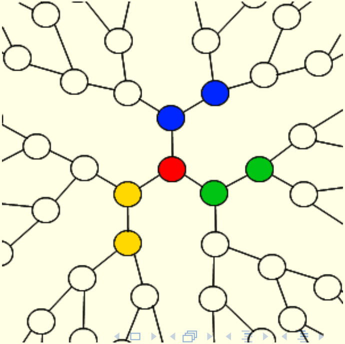
Clearly, $A_d \subseteq B_d$.

$d = 3, t = n = 7$

$(1, 1, 4), (1, 4, 1), (4, 1, 1),$

$(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1),$

$(3, 1, 2), (3, 2, 1), (2, 2, 2).$



Definition 3

$$A_d = \{(a_1, a_2, \cdots, a_d) \mid 1 \leq a_i < \frac{n}{2}, \sum_{i=1}^{d} a_i = n-1\}, \text{ and}$$

$$B_d = \{(b_1, b_2, \cdots, b_d) \mid b_i \in \mathbb{N}, \sum_{i=1}^{d} b_i = n-1\}.$$

Clearly, $A_d \subseteq B_d$. Moreover,

$$|B_d| = \binom{n-1-d+d-1}{d-1} = \binom{n-2}{d-1},$$

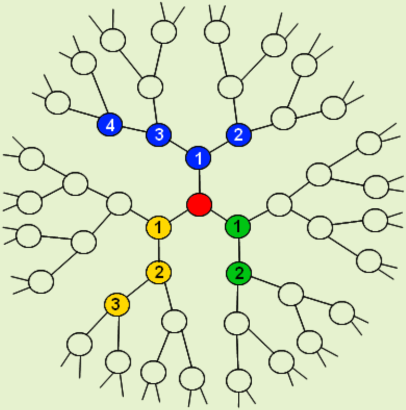
$$|A_d| = \binom{n-2}{d-1} - d \cdot \binom{\lfloor \frac{n}{2} \rfloor - 1}{d-1}.$$

$d = 3, t = n = 7$

$(1,1,4), (1,4,1), (4,1,1),$
$(1,2,3), (1,3,2), (2,1,3), (2,3,1), (3,1,2), (3,2,1), (2,2,2).$

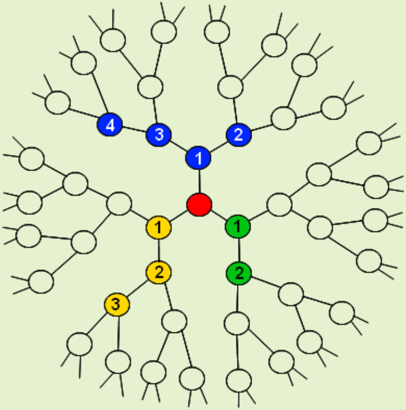
Given $(t^v_{v_1}, t^v_{v_2}, \cdots, t^v_{v_d})$, the total number of ways to spread a rumor is

$$\frac{(n-1)!}{t^v_{v_1}!t^v_{v_2}!\cdots t^v_{v_d}!} \cdot \prod_{k=1}^{d} \prod_{i=1}^{t^v_{v_k}} \left((d-2)(i-1)+1\right). \tag{2}$$



Given $(t_{v_1}^v, t_{v_2}^v, \cdots, t_{v_d}^v)$, the total number of ways to spread a rumor is

$$(n-1)! \prod_{k=1}^{d} \frac{\prod_{i=1}^{t_{v_k}^v} \left((d-2)(i-1)+1\right)}{t_{v_k}^v!}. \qquad (2)$$



Detection probability of d-regular trees

$$P_d(n) = \frac{\displaystyle\sum_{(t^v_{v_1}, t^v_{v_2}, \cdots, t^v_{v_d}) \in A_d} \left(\prod_{k=1}^{d} \frac{\prod_{k=1}^{t^v_{v_k}} \left((d-2)(i-1)+1\right)}{t^v_{v_k}!} \right)}{\displaystyle\sum_{(t^v_{v_1}, t^v_{v_2}, \cdots, t^v_{v_d}) \in B_d} \left(\prod_{k=1}^{d} \frac{\prod_{k=1}^{t^v_{v_k}} \left((d-2)(i-1)+1\right)}{t^v_{v_k}!} \right)}. \tag{3}$$

Detection probability of d-regular trees

$$P_d(n) = \frac{\displaystyle\sum_{(t^v_{v_1}, t^v_{v_2}, \cdots, t^v_{v_d}) \in A_d} \left(\prod_{k=1}^{d} \frac{\prod_{k=1}^{t^v_{v_k}} \left((d-2)(i-1)+1 \right)}{t^v_{v_k}!} \right)}{\displaystyle\sum_{(t^v_{v_1}, t^v_{v_2}, \cdots, t^v_{v_d}) \in B_d} \left(\prod_{k=1}^{d} \frac{\prod_{k=1}^{t^v_{v_k}} \left((d-2)(i-1)+1 \right)}{t^v_{v_k}!} \right)}. \tag{3}$$

Theorem 3

If G is a **3-regular tree**, then we have that

$$\lim_{t \to \infty} \mathrm{P}(E_t) = \frac{1}{4}.$$

Let

- $w_{dn} = \prod_{i=1}^{n} (d-2)(i-1) + 1.$

Let

- $w_{dn} = \prod_{i=1}^{n} (d-2)(i-1)+1.$
- $f(x)$: The exponential generating function for the sequence $\{w_{dn}\}_{n=1}^{\infty}$

$$f(x) = \sum_{n=1}^{\infty} \frac{w_{dn}}{n!} x^n = \sum_{n=1}^{\infty} \frac{\prod_{i=1}^{n}(d-2)(i-1)+1}{n!} x^n. \tag{4}$$

Let

- $w_{dn} = \prod_{i=1}^{n} (d-2)(i-1)+1$.
- $f(x)$: The exponential generating function for the sequence $\{w_{dn}\}_{n=1}^{\infty}$

$$f(x) = \sum_{n=1}^{\infty} \frac{w_{dn}}{n!} x^n = \sum_{n=1}^{\infty} \frac{\prod_{i=1}^{n}(d-2)(i-1)+1}{n!} x^n. \tag{4}$$

And we have

- $(1-ax)^{-\frac{1}{a}} = \sum_{n=0}^{\infty} \binom{-\frac{1}{a}}{n} (-ax)^n = 1 + \sum_{n=1}^{\infty} \frac{\prod_{i=1}^{n} a(i-1)+1}{n!} x^n.$

Let

- $w_{dn} = \prod_{i=1}^{n} (d-2)(i-1)+1.$
- $f(x)$: The exponential generating function for the sequence $\{w_{dn}\}_{n=1}^{\infty}$

$$f(x) = \sum_{n=1}^{\infty} \frac{w_{dn}}{n!} x^n = \sum_{n=1}^{\infty} \frac{\prod_{i=1}^{n}(d-2)(i-1)+1}{n!} x^n. \tag{4}$$

And we have

- $(1-ax)^{-\frac{1}{a}} = \sum_{n=0}^{\infty} \binom{-\frac{1}{a}}{n} (-ax)^n = 1 + \sum_{n=1}^{\infty} \frac{\prod_{i=1}^{n} a(i-1)+1}{n!} x^n.$

Hence we immediately know $f(x) = (1-ax)^{-\frac{1}{a}} - 1$ where $a = d-2$.

Detection probability of d-regular trees

Let $a = d - 2$,

$$P_d(n) = 1 - \frac{d \sum_{m \geq \frac{n}{2}}^{n-d} \frac{w_{dm}}{m!} \left([x^{n-1-m}] F_{d-1}(x) \right)}{[x^{n-1}] F_d(x)}. \quad (5)$$

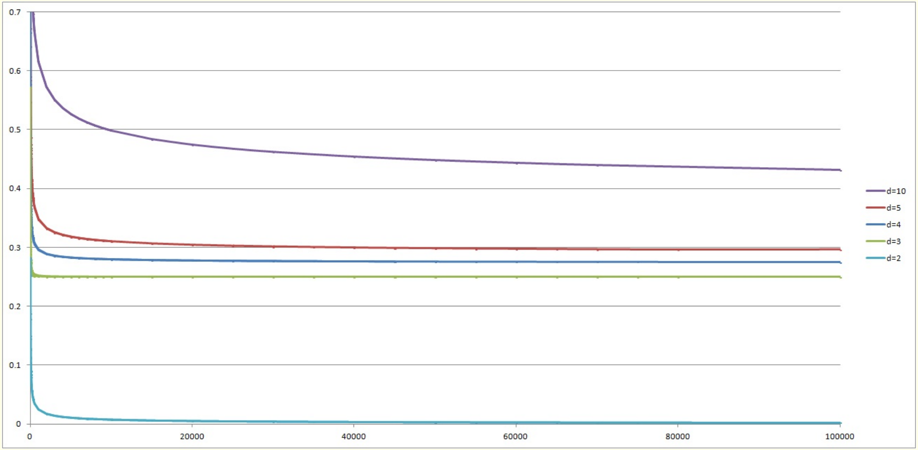
Where

$$w_{dn} = \prod_{i=1}^{n} \left(a\left(i - 1\right) + 1 \right),$$

$$[x^n] F_k(x) = \sum_{l=1}^{k} (-1)^{k-l} \binom{k}{l} \frac{\prod_{i=1}^{n} \left(l + a(i-1) \right)}{n!}.$$

Conclusion

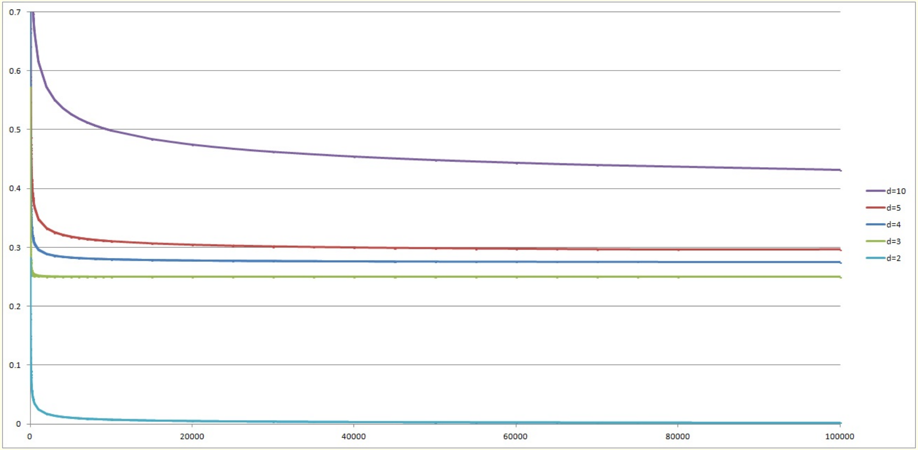
General behavior of detection probability for several d's and $n \leq 100,000$.



Conclusion

General behavior of detection probability for several d's and $n \leq 100,000$.

- $\frac{1}{2} \geq P_d(n) \geq P_{d'}(n) \geq \frac{1}{4}$ if $d \geq d' \geq 3$



Reference

- N. T. J. Bailey. The Mathematical Theory of Infectious Diseases and its Applications. Griffin, London, 1975.
- A. Ganesh, L. Massoulie, and D. Towsley. The effect of network topology on the spread of epidemics. Proc. IEEE INFOCOM, 2:1455-1466, 2005.
- C. Moore and M. E. J. Newman. Epidemics and percolation in in small-world networks. Phys. Rev. E, 61:5678-5682, 2000.
- M. E. J. Newman. The spread of epidemic disease on networks. 2002.
- R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. Phys. Rev. Lett., 86:3200-3203, 2001.
- D. Shah and T. Zaman, Rumors in a Network: Who's the Culprit?, IEEE Transactions on Information Theory, vol. 57, pp. 5163-5181, August 2011.
- J. H. van Lint and R. M. Wilson. *A Course in Combinatorics*, Cambridge University Press, Cambridge, second edition, 2001.
- D.B. West, *Introduction to Graph Theory* (Prentice-Hall, Upper Saddle River, NJ, 1996).

Thank you for listening.

祝 張鎮華教授

六十大壽

生日快樂