

삼각형 구조를 갖는 네트워크와  
임의의 접촉 패턴을 고려한  
정보 확산의 상전이 현상 연구

임성수

Applied Algorithm Lab, KAIST  
(joint work with 곽남주, 정교민)

## ■ 네트워크 분석

- 구성원들 사이의 상호 작용 및 연결 구조를 표현
- 네트워크 분석을 통한 현상의 이해 및 예측

## ■ 상전이 현상

- 정보 확산이 급격히 발생하는 조건(tipping point) 존재
- 정보가 광범위하게 확산되는 경우를 분석

## ■ 광범위 확산

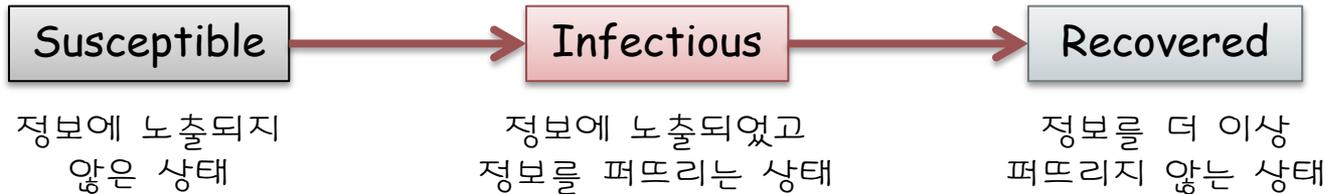
- 정보의 광범위 확산의 발생 조건
- 광범위 확산이 발생할 확률
- 광범위 확산이 발생시 확산 규모



# SIR 모형

- **SIR 모형** (Kermack and McKendrick 1927)

- 정보 확산 과정을 모델링



- 초기 모형

- 연결된 두 구성원들 사이에서 정보가 확산될 확률이 항상 동일
- Complete graph, Random graph
- Homogeneous contact rates

- 최근 연구 결과

- Generalized random graph, Configuration model (Newman 2002)
- **Random clustered network** (Newman 2009, Miller 2009)
- **Heterogeneous contact rates** (Miller 2007, Kenah and Robins 2007)

# 주요 결과

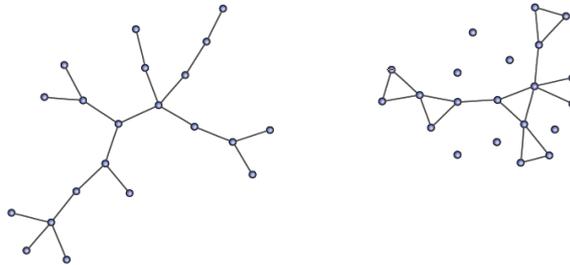
- 본 연구에서 고려한 모형
  - Random clustered network
    - 연결수와 삼각형 구조의 분포를 고려한 일반화된 무작위 그래프
  - Arbitrary contact rate
    - 두 구성원 사이에서 정보가 확산될 가능성이 임의로 주어진 경우까지 분석
- 주요 결과
  - 위 모형에 대해 광범위 확산의 발생 확률 및 규모를 계산 가능!
  - 삼각형 구조가 광범위 확산의 발생 조건에 영향을 끼친다는 사실을 규명!



두 구성원 사이에 연결이 존재한다고  
항상 정보가 전파되는 것은 아니다

## ■ Random clustered network

- 연결 상태를 모를 때 실제 네트워크의 성질을 반영하는 네트워크 생성
- 각 구성원의 평균 연결수:  $d_1, d_2, \dots, d_n$
- 각 구성원이 속하는 삼각형 수:  $t_1, t_2, \dots, t_n$
- 이전의 모형들은 삼각형 구조를 고려하지 못함
  - Random graph ( $d_1 = d_2 = \dots = d_n, t_1 = t_2 = \dots = t_n = 0$ )
  - Generalized random graph ( $t_1 = t_2 = \dots = t_n = 0$ )



## ■ Arbitrary contact rates

- $f(i, j)$ : 구성원  $i$ 가 구성원  $j$ 에게 정보를 전파할 확률
- E.g.  $f(i, j) = \frac{c}{d_i}$  or  $\frac{c}{d_j}$ ,  $c$ 는 상수, 또는 멱함수 분포를 따름

# 주요 결과

- 광범위 확산의 발생 확률( $P$ ) 및 규모( $S$ )
  - 각 구성원  $i$ 에 대해 광범위 확산을 이끌어 낼 확률  $P_i$  계산
  - 각 구성원  $j$ 에 대해 확산에 속하게 될 확률  $S_j$  계산
  - 삼각형 구조를 고려한 Branching process와 Taylor series approximation을 통해  $P_i, S_j$  계산 ( $k(i, j)$ :  $i, j$ 에 연관된 함수)

$$P \approx 1 - \frac{1}{n} \sum_i \exp \left\{ - \sum_j k(i, j) P_j \right\}$$
$$S \approx 1 - \frac{1}{n} \sum_j \exp \left\{ - \sum_i k(j, i) S_i \right\}$$

- 광범위 확산이 발생 조건을 규명
  - Brower's fixed point theorem과 partial differential equation을 통해 광범위 확산 발생 조건에 대한 lower/upper bound 계산

# 삼각형을 고려하지 않을 때

- THM 1: 평균 연결수가  $d_1, \dots, d_n$ 을 따르는 generalized random network의 경우 구성원  $i$ 가 광범위 확산을 이끌어 낼 확률  $P_i$ 와 구성원  $j$ 가 광범위 확산이 포함될 확률  $S_j$ 는 다음을 만족한다.

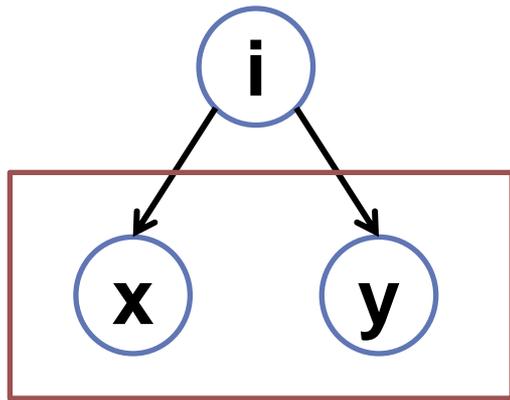
$$P_i = 1 - \exp \left\{ - \frac{d_i}{\sum_x d_x} \sum_j d_j f(i, j) P_j \right\}$$
$$S_j = 1 - \exp \left\{ - \frac{d_j}{\sum_x d_x} \sum_i d_i f(j, i) S_i \right\}$$

- THM 2: 위 조건 하에서 광범위 확산이 발생할 확률  $P$ 와 발생시 평균 규모  $S$ 는 다음을 만족한다.

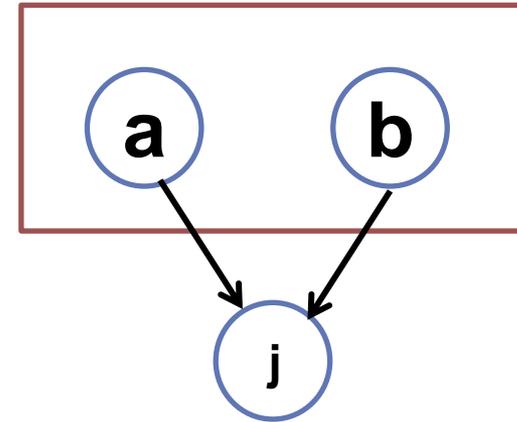
$$P = \frac{1}{n} \sum_i P_i, \quad S = \frac{1}{n} \sum_j S_j$$

# 삼각형을 고려하지 않을 때

- THM 1 증명 아이디어: 가지치기 과정과 그 역 과정을 응용



i가 광범위 확산이 이끌지 않는 경우  
= i에게 정보를 받은 구성원들이  
광범위 확산을 이끌지 않는 경우



j가 광범위 확산에 속하지 않는 경우  
= j에게 정보를 전달한 구성원들이  
광범위 확산에 속하지 않는 경우

- 확률적 과정을 모델링하여 광범위 확산 발생 확률과 규모를 예측!

# 삼각형을 고려할 때

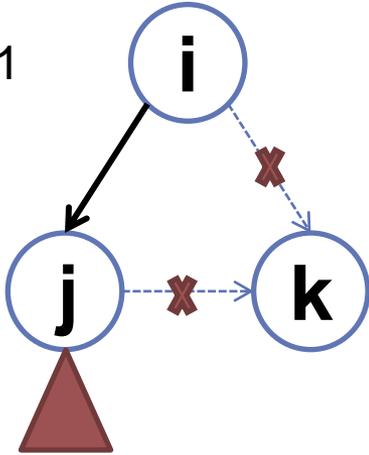
- THM 3: 평균 연결수가  $d_1, \dots, d_n$ , 평균 삼각형 연결수가  $t_1, \dots, t_n$ 을 따르는 random clustered network의  $P_i$ 와  $S_j$ 는 다음을 만족한다.

$$P_i = 1 - \exp \left\{ - \sum_j k(i, j) P_j \right\}$$
$$S_j = 1 - \exp \left\{ - \sum_i k(j, i) S_i \right\}$$

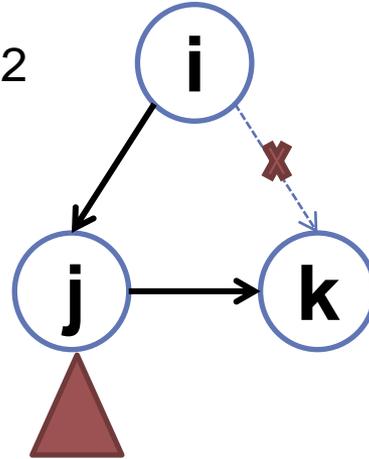
- THM 1은 THM 3의 특별한 경우 ( $t_1, \dots, t_n$ )
- 삼각형 구조를 함께 고려한 모형을 사용하여 더 정확하게 분석해본다!
- 가지치기 과정의 일반화를 사용 (다음쪽)

# 삼각형을 고려할 때

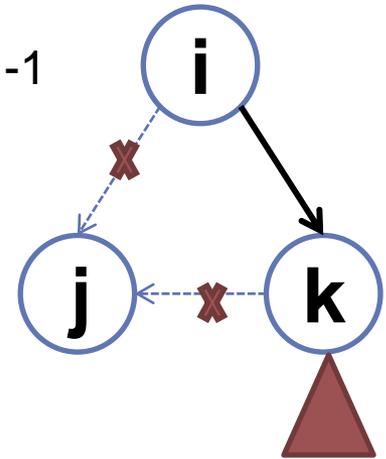
Case I-1



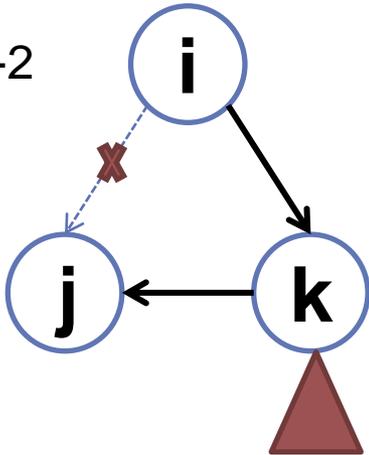
Case I-2



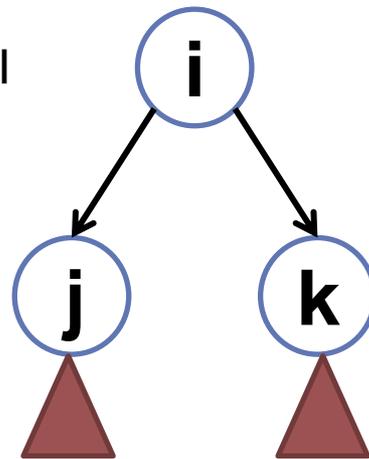
Case II-1



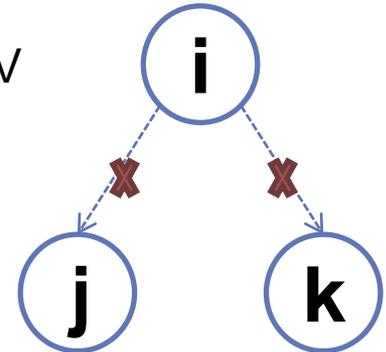
Case II-2



Case III

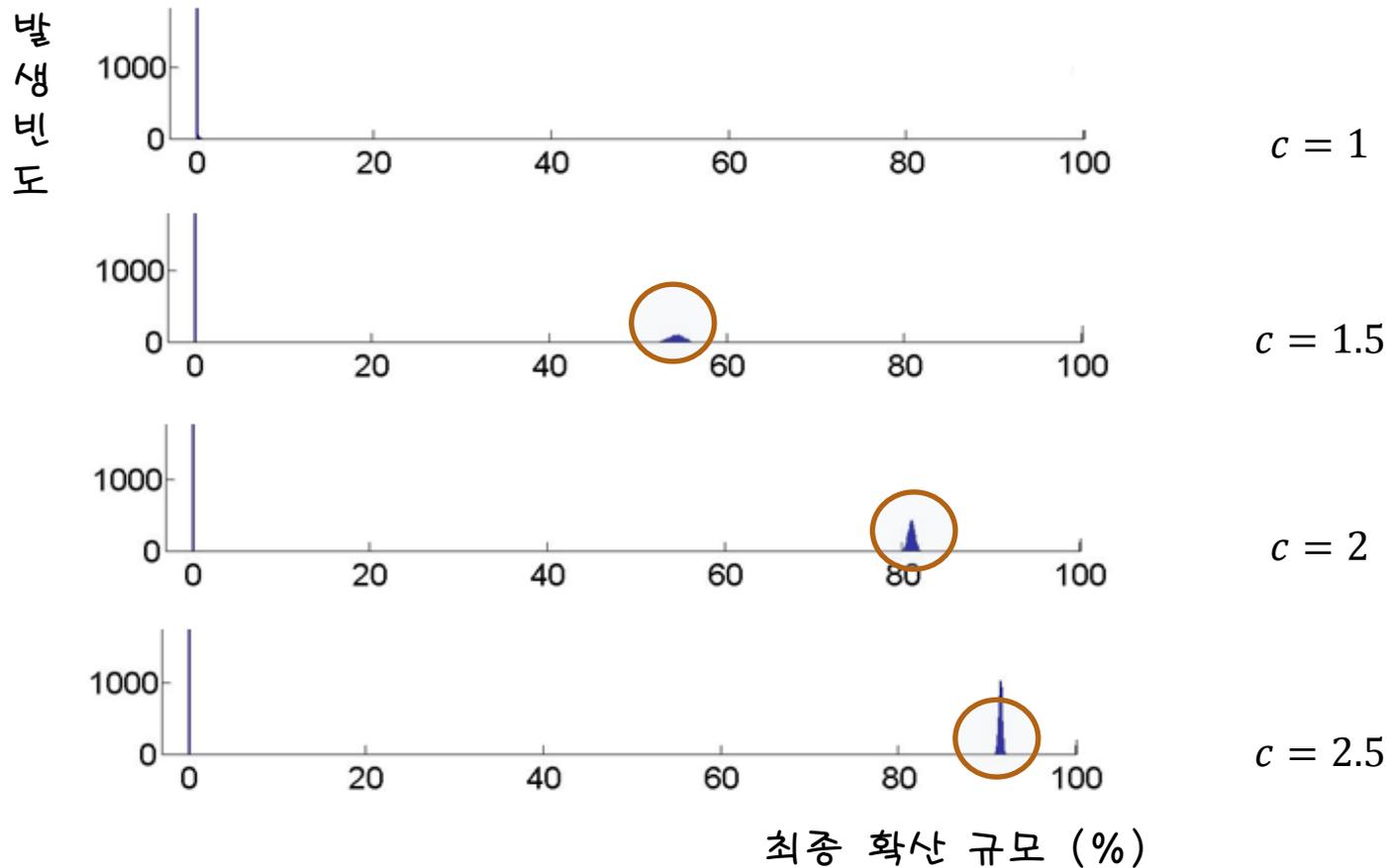


Case IV



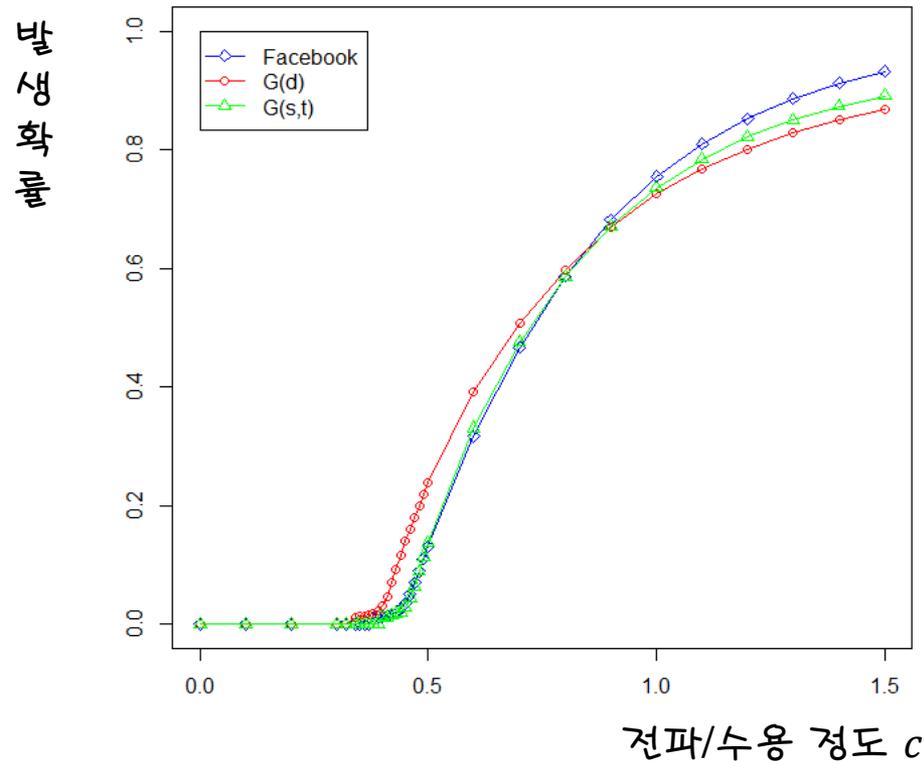
# 실험 결과

- 광범위 확산 발생 확률, 규모 모두 특정 값 주변에 집중  
(예:  $f(i,j) = \frac{c}{d_j}$ , Facebook,  $n = 63K$ )



# 실험 결과

- Random clustered network를 통해 계산한 광범위 확산 발생 확률. 규모가 실제 네트워크와 더 유사한 결과를 보임  
(예:  $f(i,j) = \frac{c}{d_i} + \frac{c}{d_j}$ , Facebook,  $n = 63K$ )



# 감사합니다

