

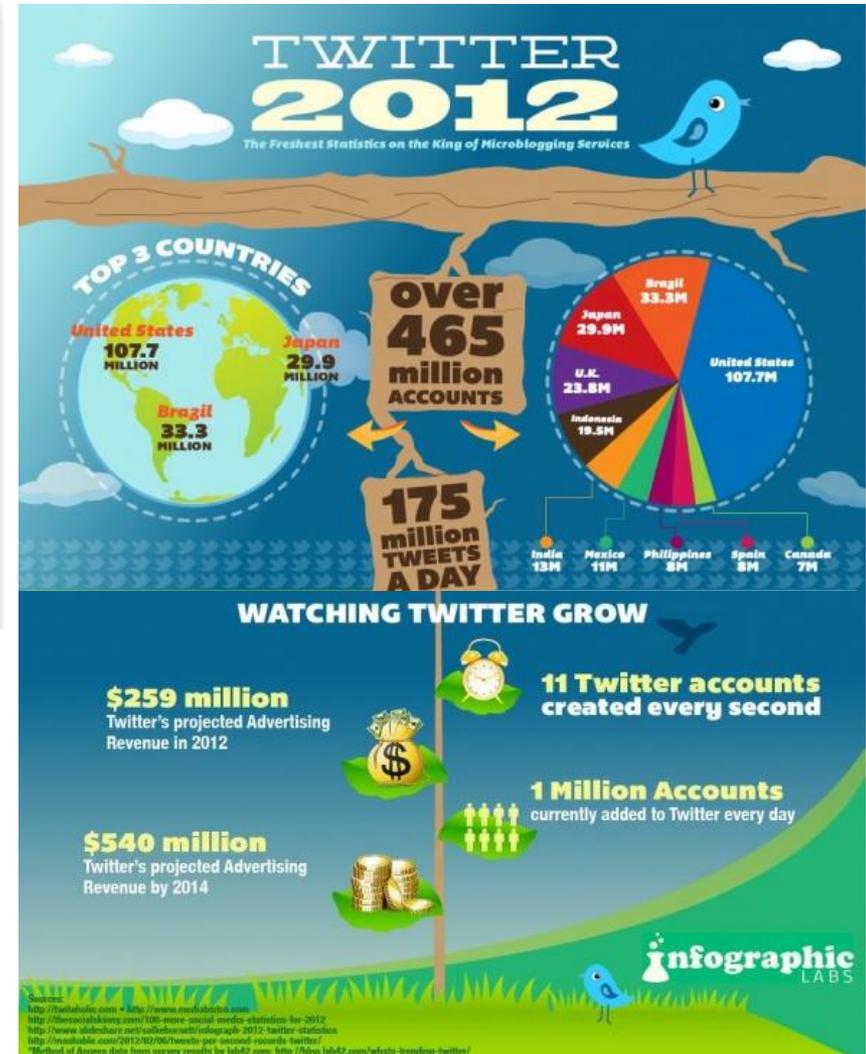
# Future Internet for SNS

경희대 컴퓨터공학과 콜로퀴엄  
2012년 6월 1일

문수복  
전산학과 /  
웹 사이언스 공학 전공

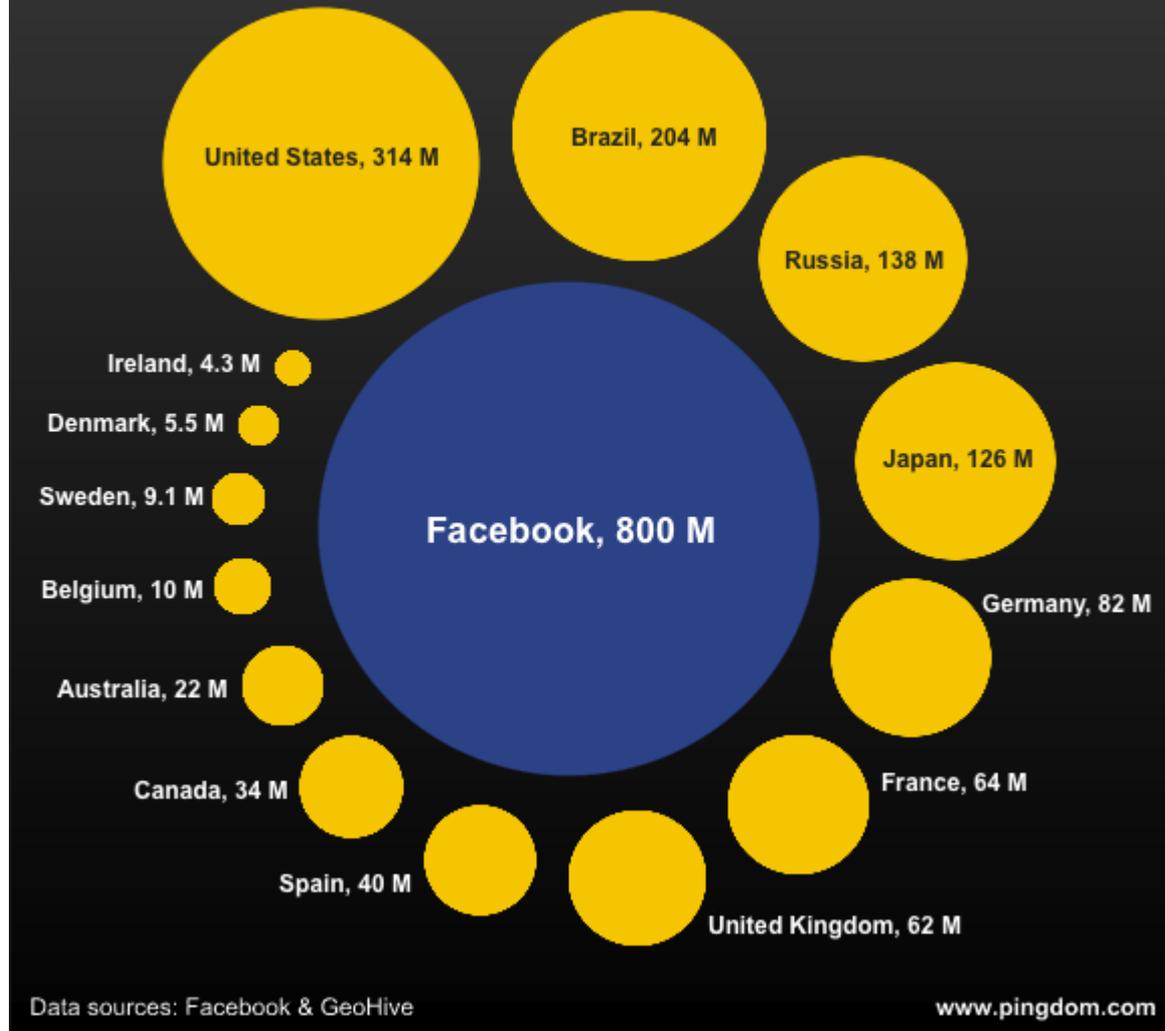
1. 연구 배경
2. 연구 목표
3. 연구 세부분야
4. 연구책임자 실적
5. 협업 대상
6. 기대효과

# 온라인 미디어의 현재



# 국가별 인구수와 비교

The Facebook user base vs. country populations



# 다양한 온라인 서비스들: 온라인-오프라인의 결합

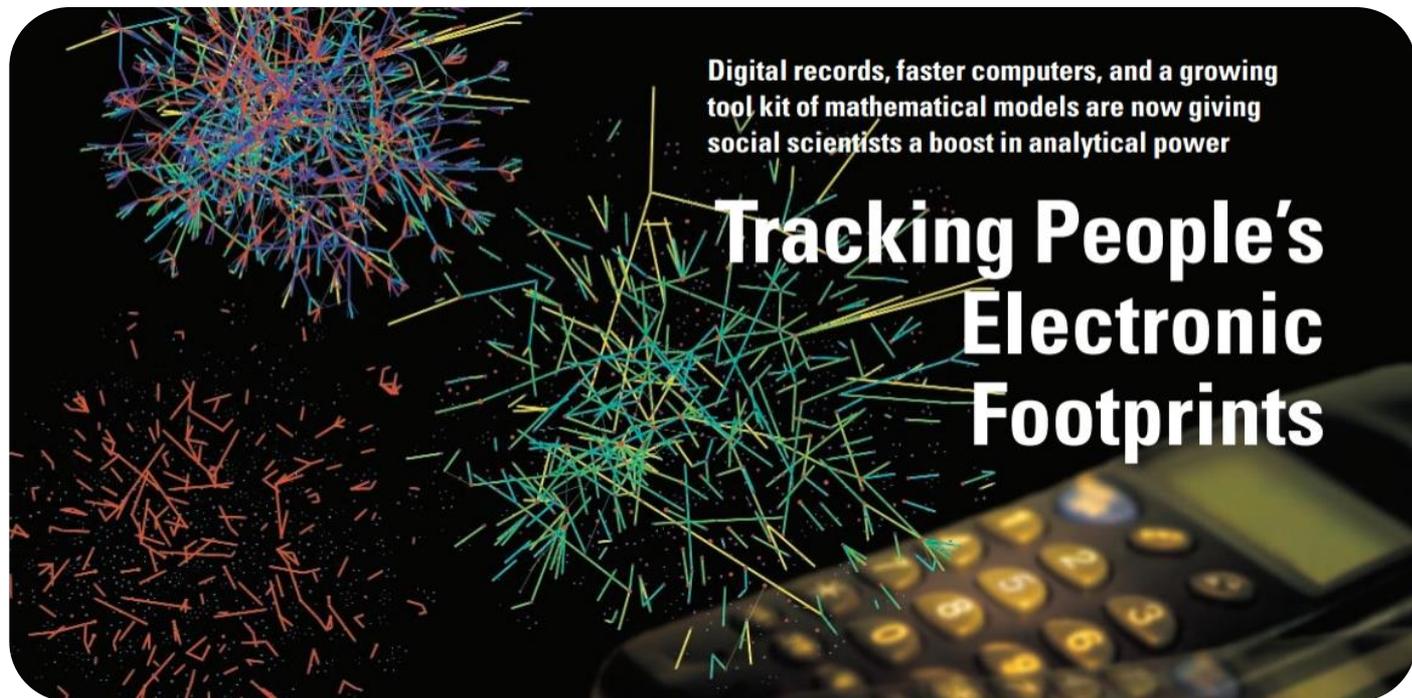


©2010 SocialGrow

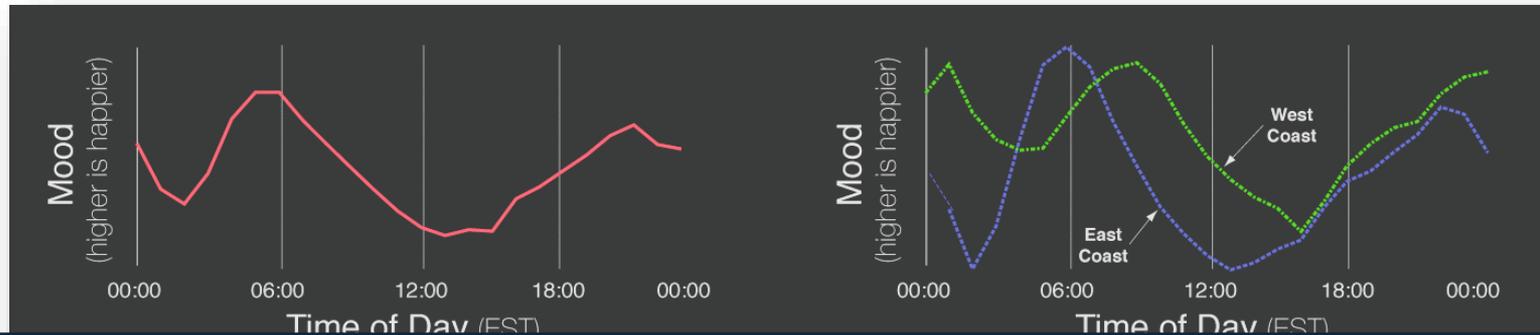
## Science Featured Articles

### Tracking People's Electronic Footprints

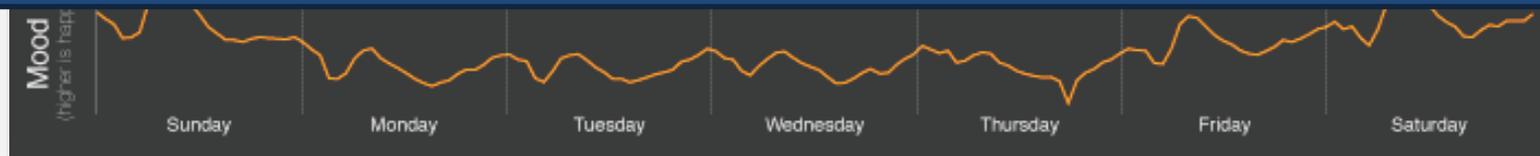
“다양한 온라인 서비스를 사용하는 사용자의 행동이 전자 로그 형태로 기록되고 있다” (Science, Nov. 2007)



# 트윗에 사용된 단어로부터 사용자의 행복도를 분석



트윗 분석을 바탕으로 지역별/시간별 행복도를 정량적으로 측정,  
국가 복지 정책에 활용 및 다양한 대상에 대한 선호도 분석 기술로 확장



<http://www.ccs.neu.edu/home/amislove/twittermood/>

# 페이스북 프로필에 기록된 연애 시작/끝의 시계열적 특성 분석



<http://www.facebook.com/notes/facebook-data-team/the-right-time-for-love-tracking-the-seasonality-of-relationship-formation/10150643989093859/>

# 언제, 어디서, 무엇을?

CHECK-INS BY CATEGORY THROUGHOUT THE



FOOD



WORK / OFFICE



SHOPS



TRAVEL



NIGHTLIFE

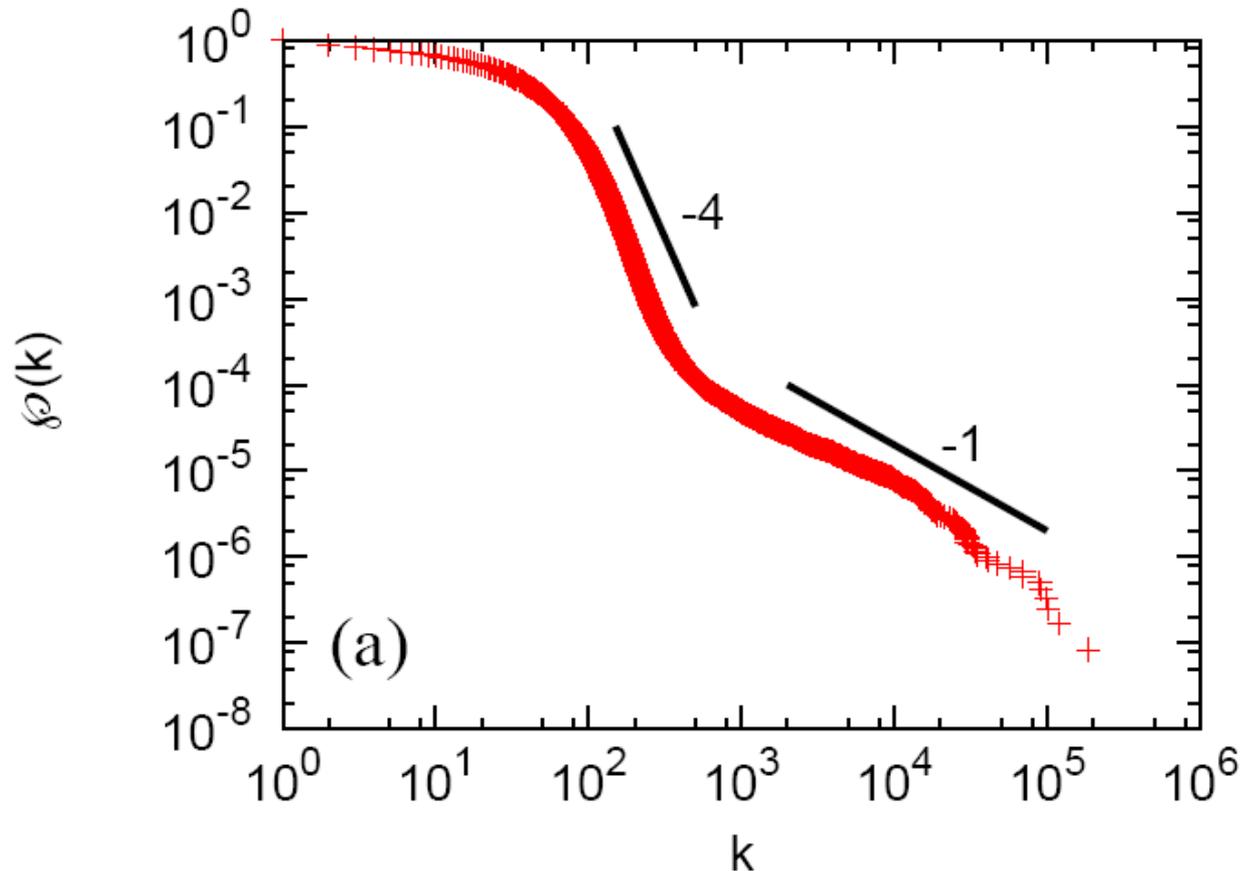
모바일 단말을 통해 사용자들의 위치 정보를 수집하여 인구 유동을 파악하고 이를 대중교통 체계 개선 및 지역기반 서비스에 활용



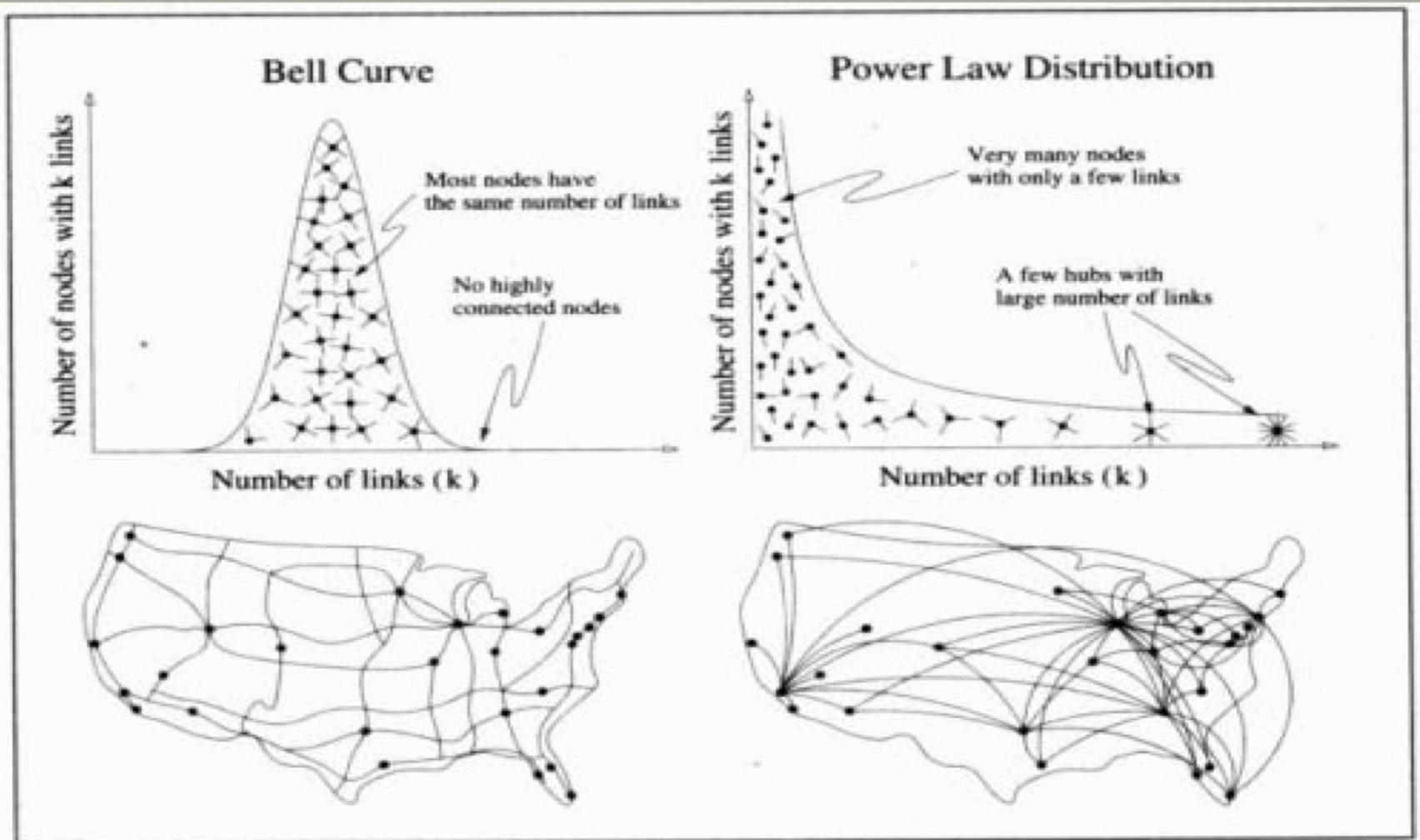
<https://ko.foursquare.com/infographics/2010infographic>

# 싸이월드 분석, 무엇부터?

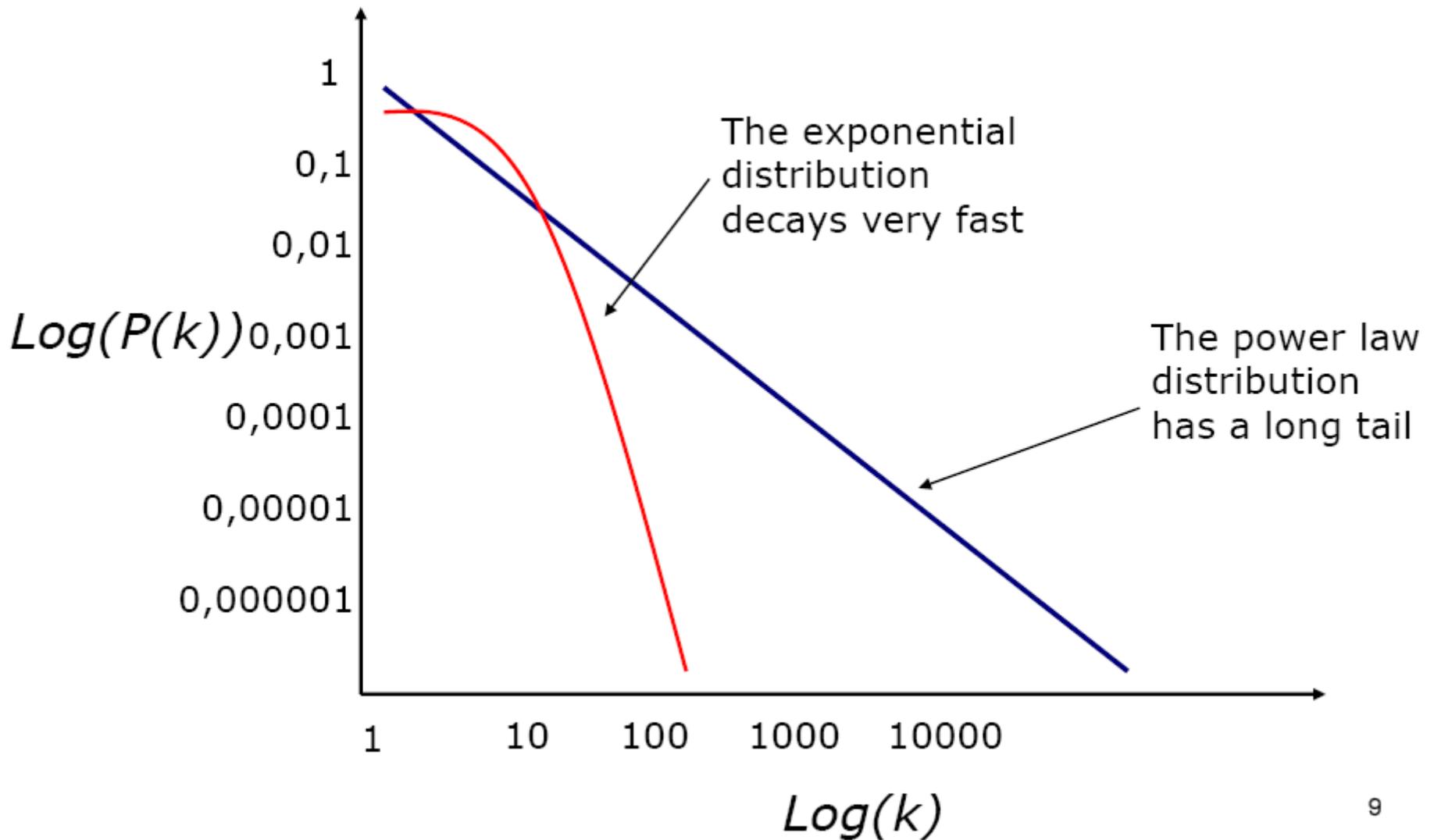
- 친구가 몇 명이나 있을까?
  - 평균?
  - 미디언?
  - 모드?
  - 10명 이상인



# Power-Law 또는 멱함수?



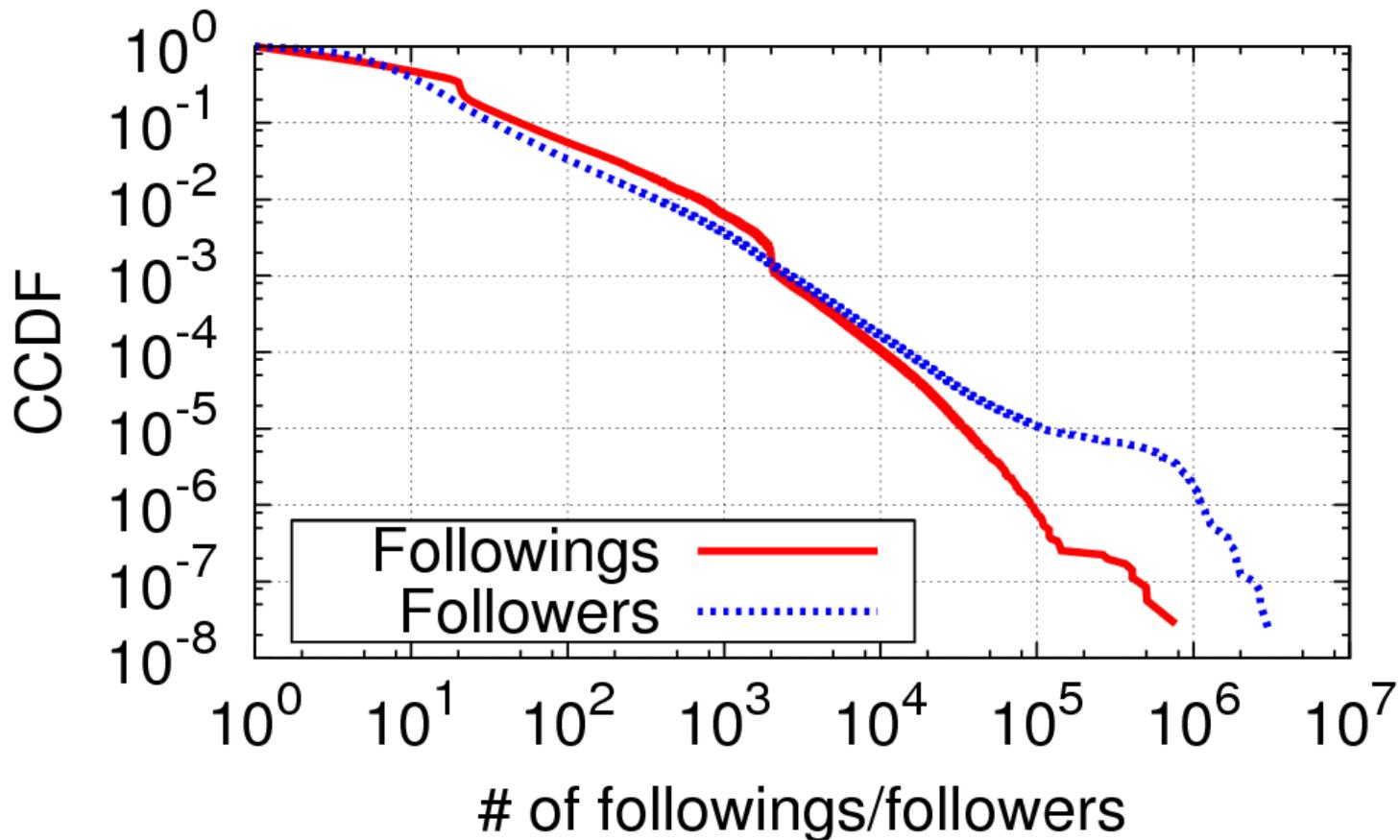
# 멱함수와 지수함수의 차이



## 사이월드를 살펴보니

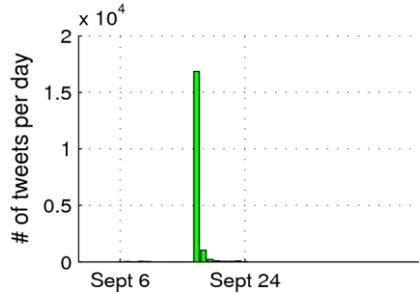
- 멱함수로 예측한 것보다 꼬리가 더 무겁다
- 왜?
  - 일반적인 인간 관계가 아닌 온라인/싸이버 관계 출현

# 그렇다면 트위터는?

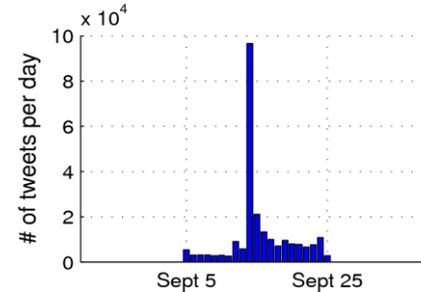


# 주로 다루는 얘기는?

**31.5%**  
"ephemeral"



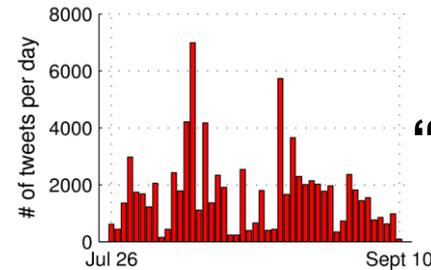
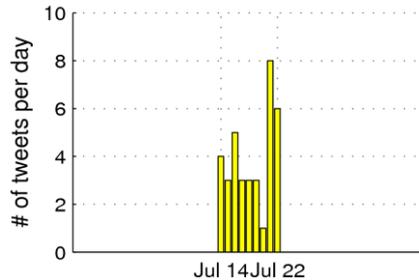
**54.3%**  
"headline news"



"backintheday"

"beyonce"

**6.9%**

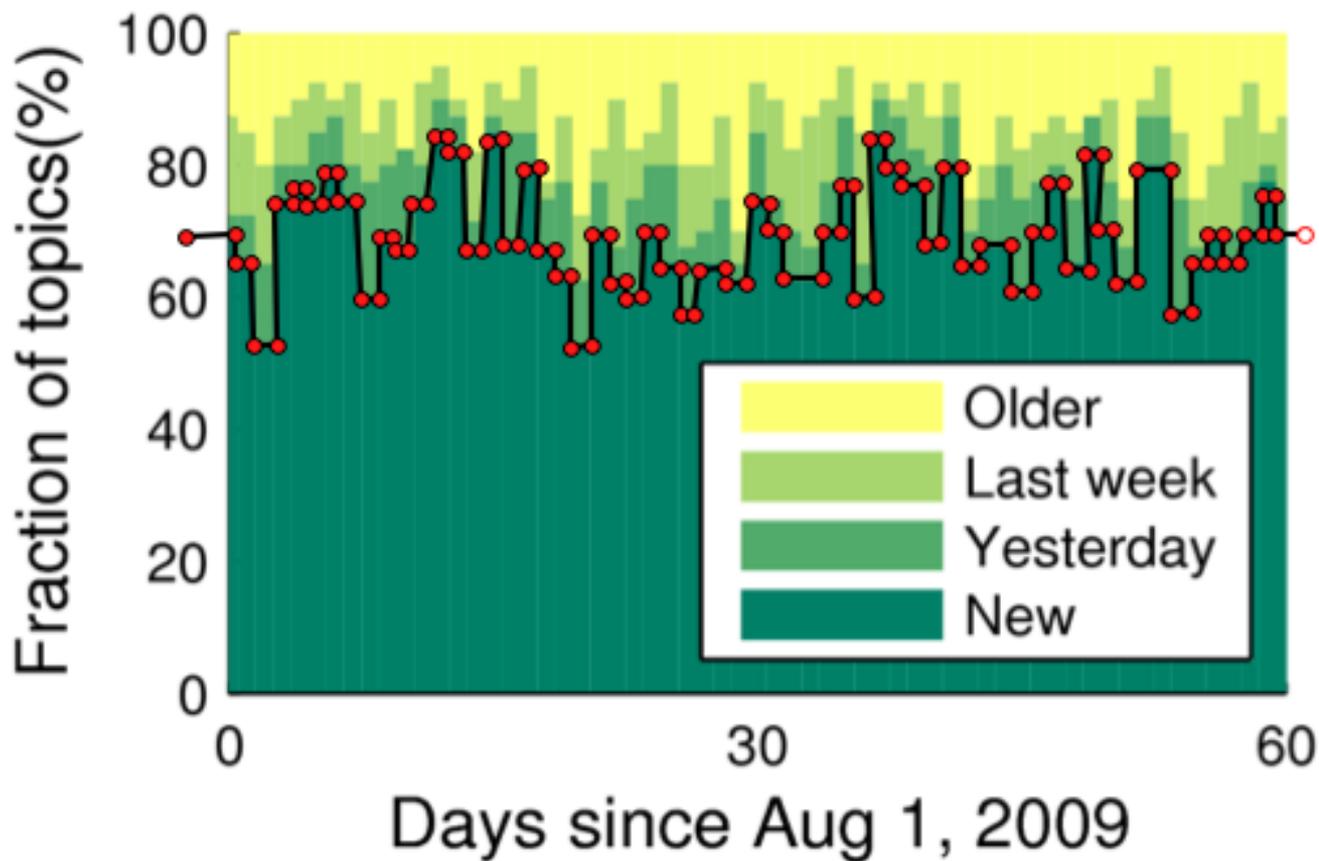


**7.3%**  
"persistent news"

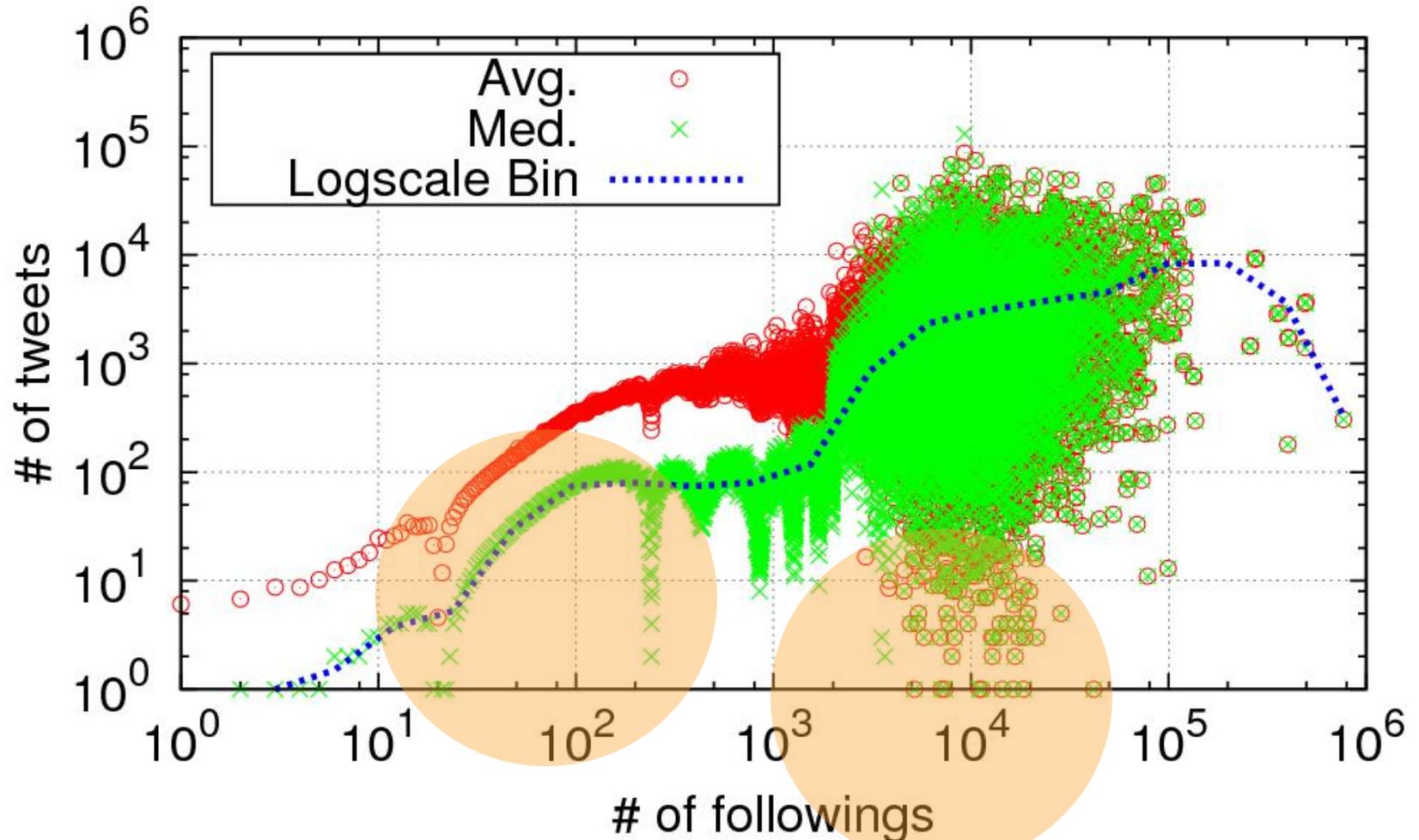
"lynn harris"

"#redsox"

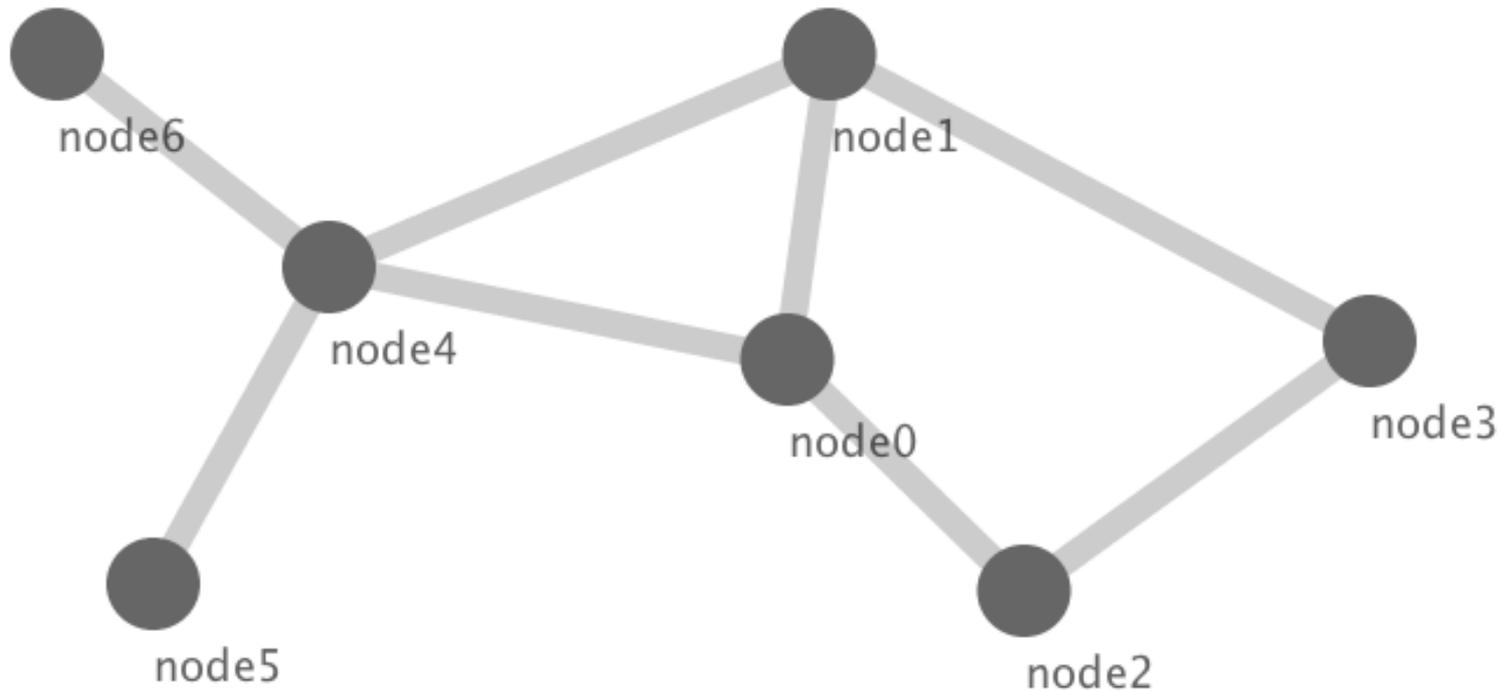
# 다루는 내용이 자주 바뀌는지?



# 팔로워가 많으면 더 열심히 트윗한다?

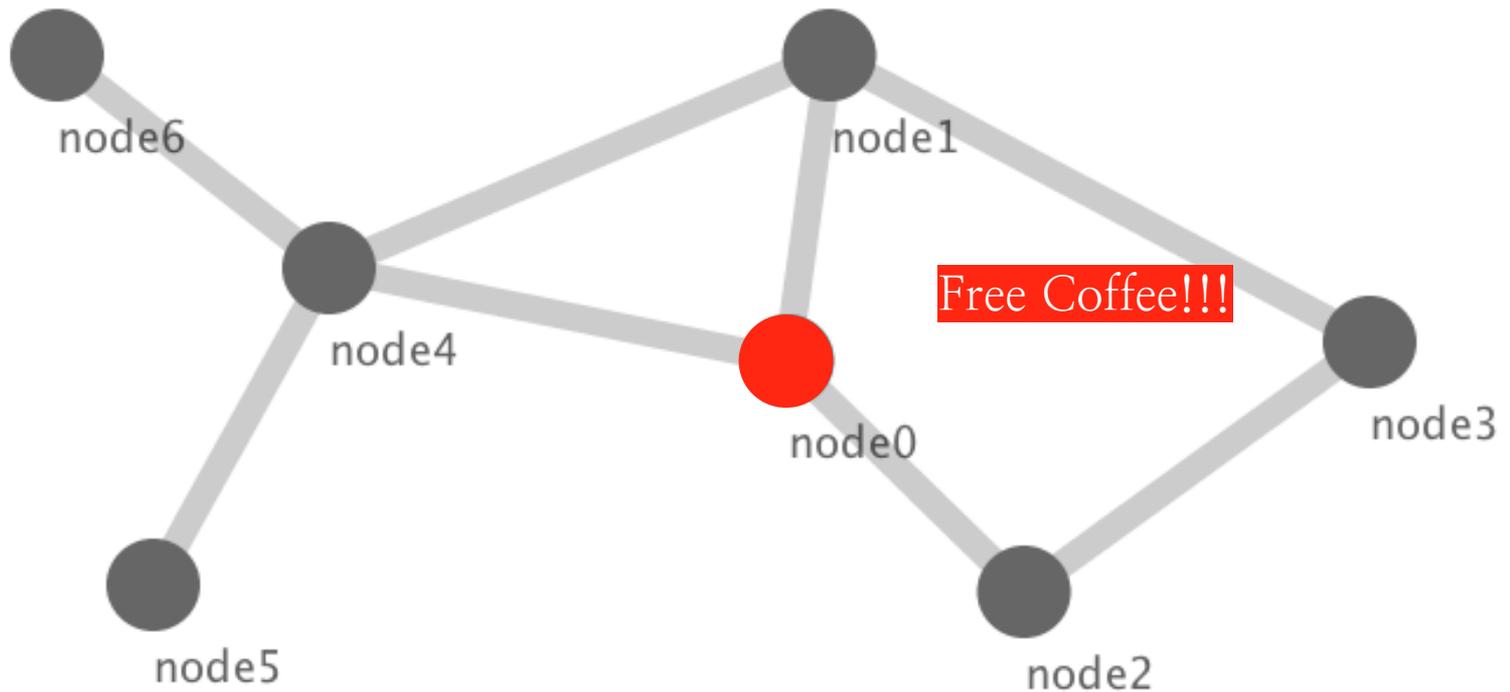


# 리트윗 효과가 얼마나 있는지?



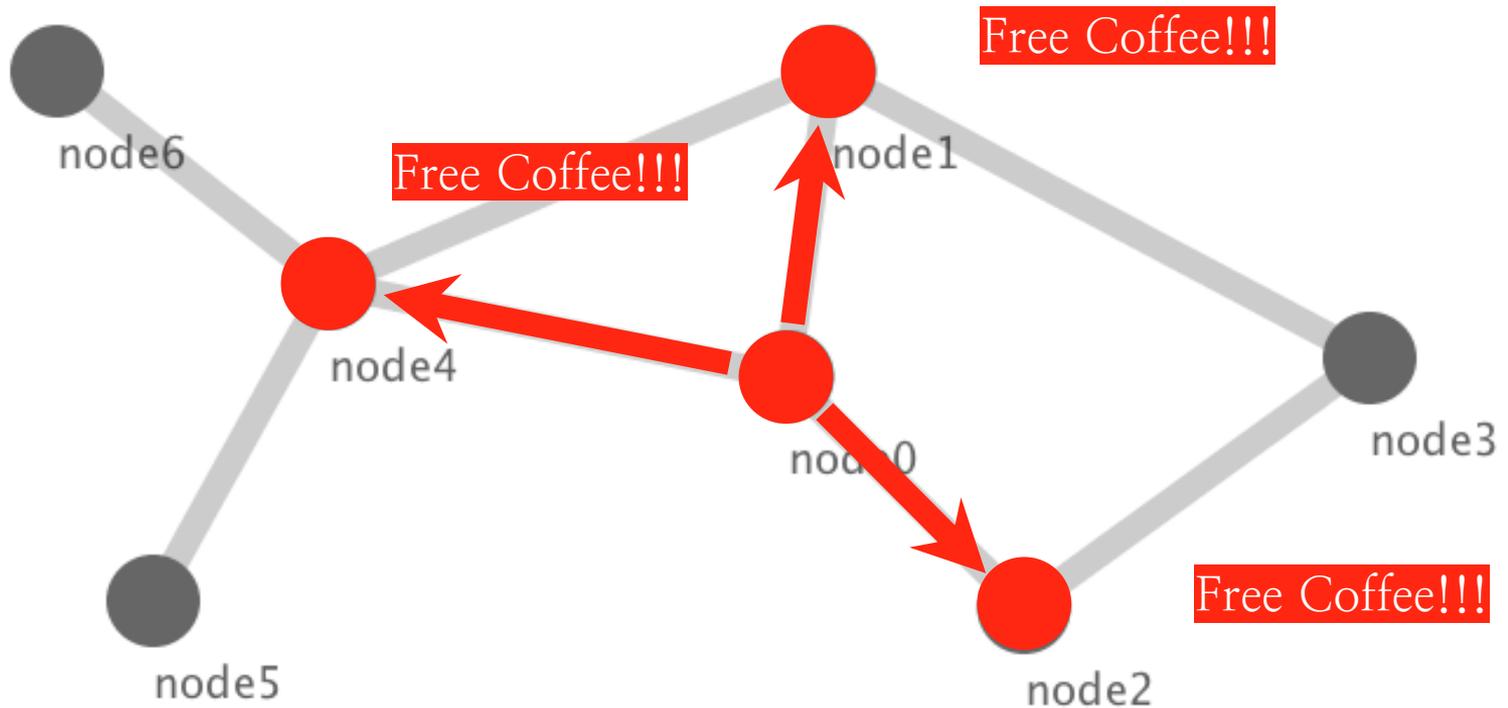
# Retweet (RT)

- Relay tweets from a following to followers



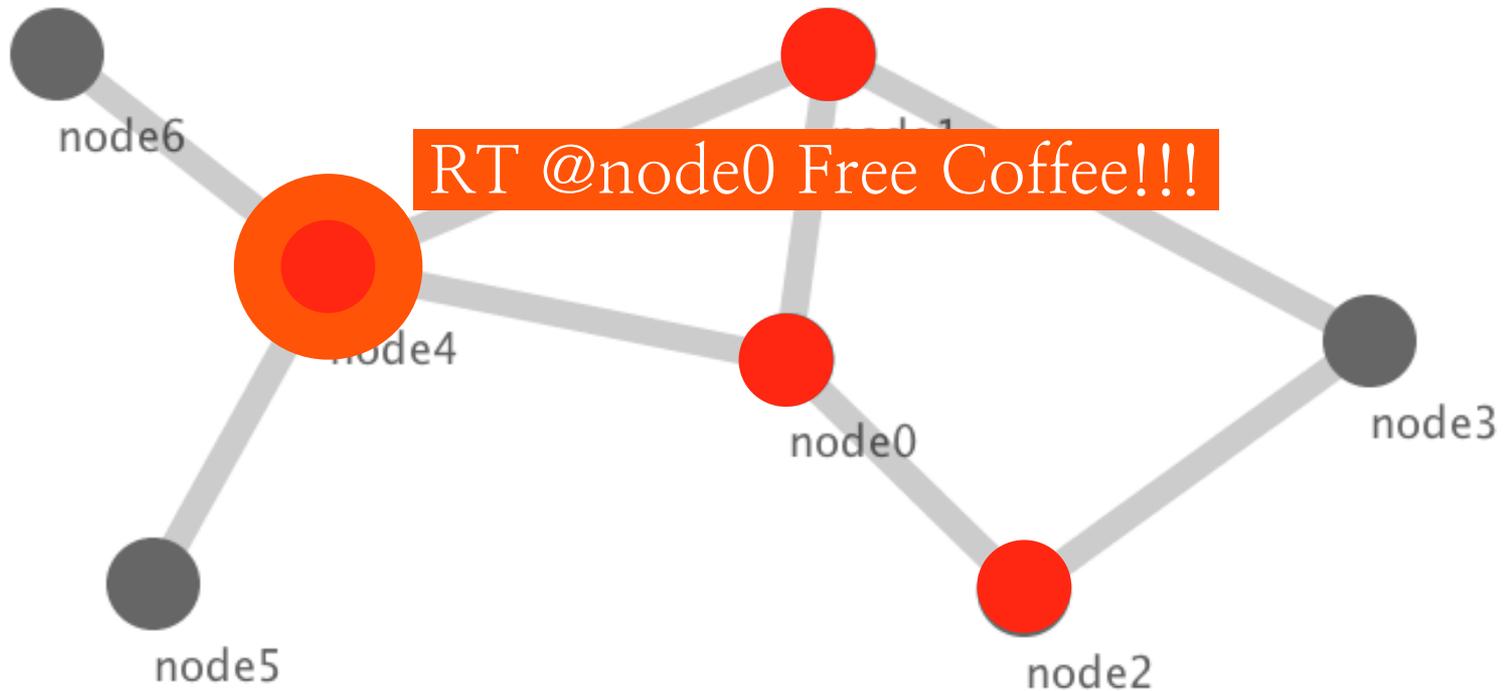
# Retweet (RT)

- Relay tweets from a following to followers



# Retweet (RT)

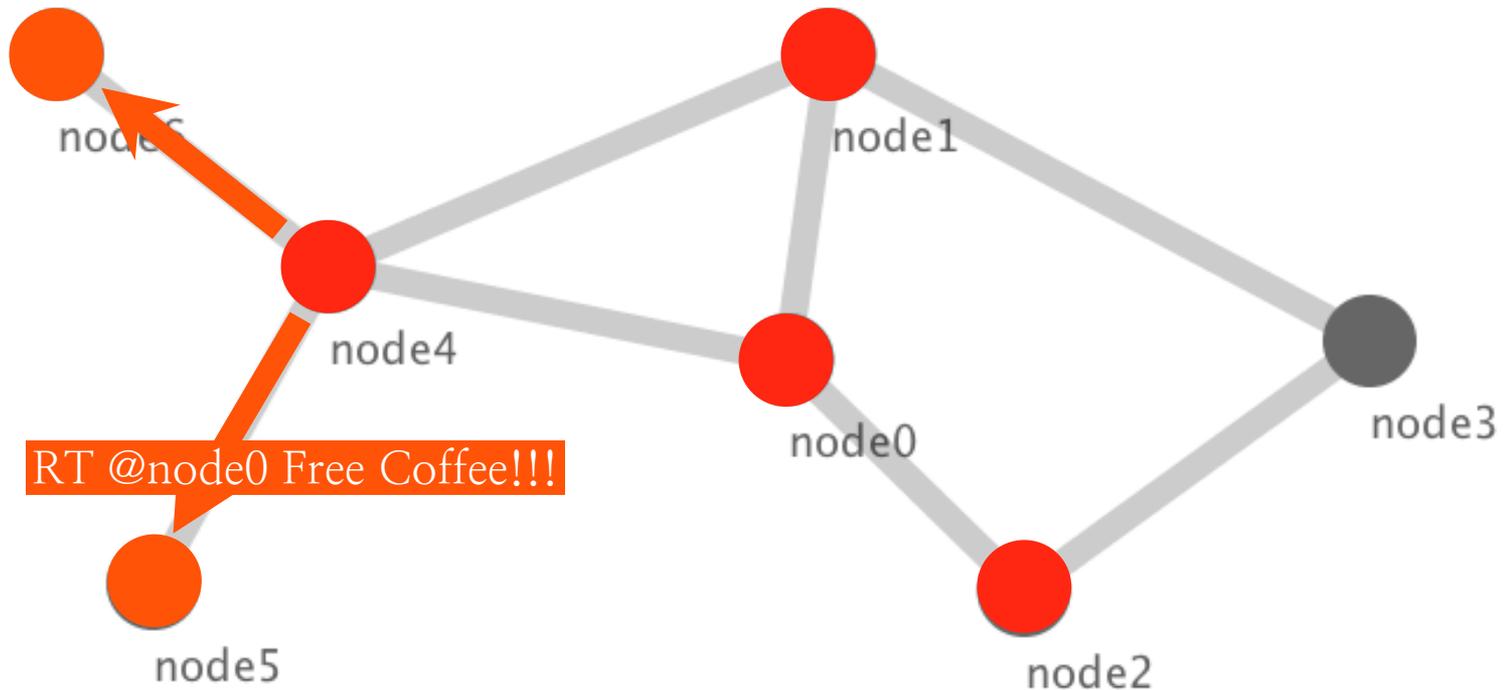
- Relay tweets from a following to followers



# Retweet (RT)

- Relay tweets from a following to followers

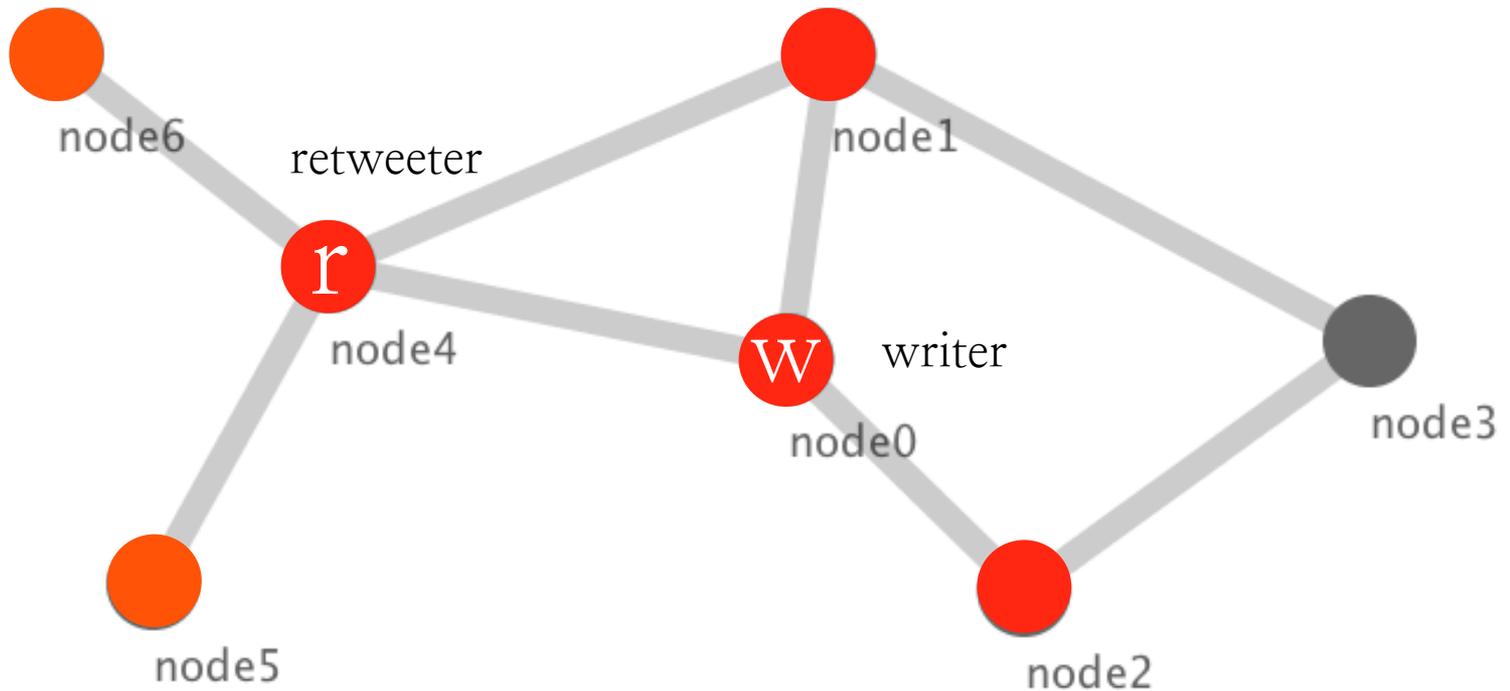
RT @node0 Free Coffee!!!



RT @node0 Free Coffee!!!

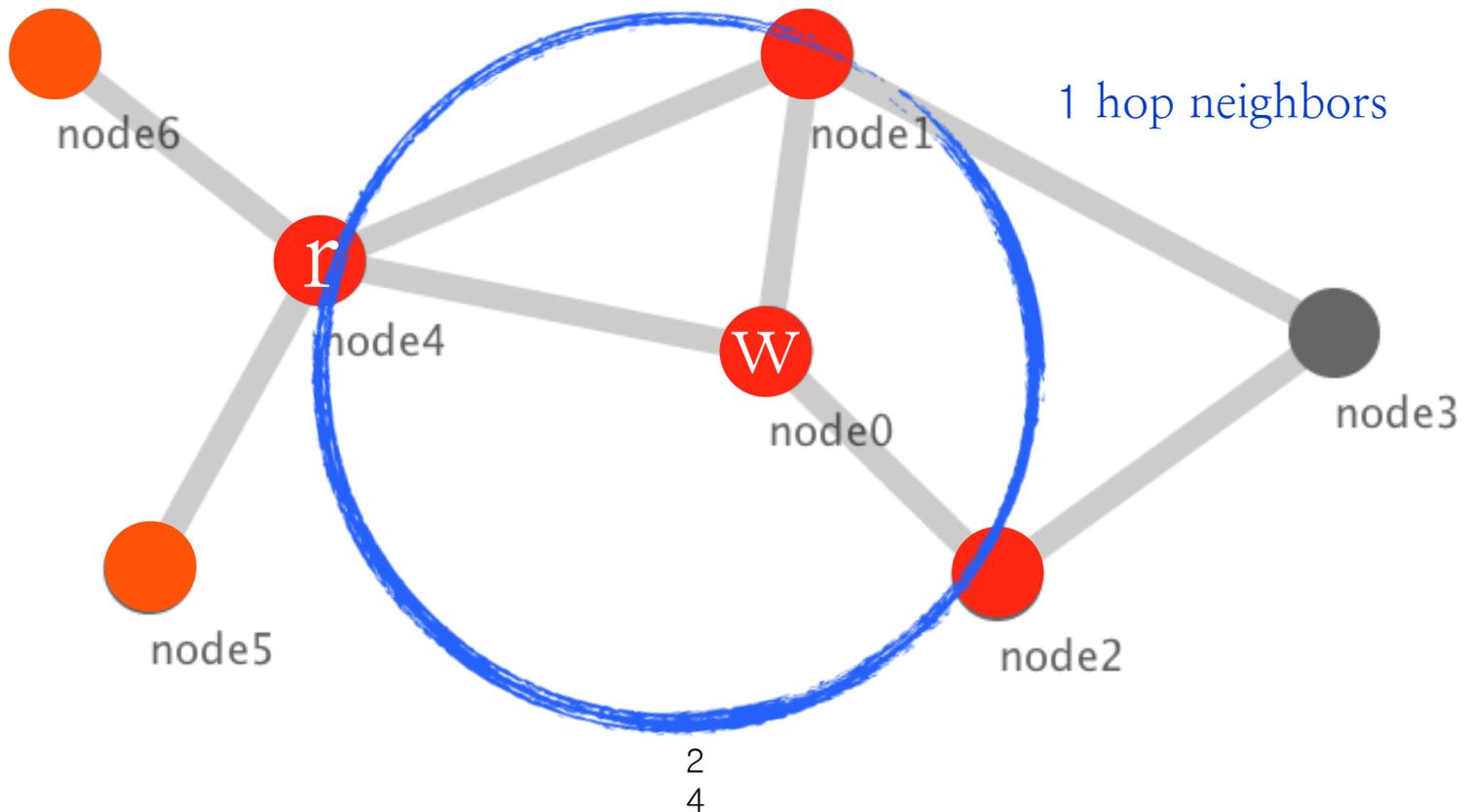
# Retweet (RT)

- Relay tweets from a following to followers

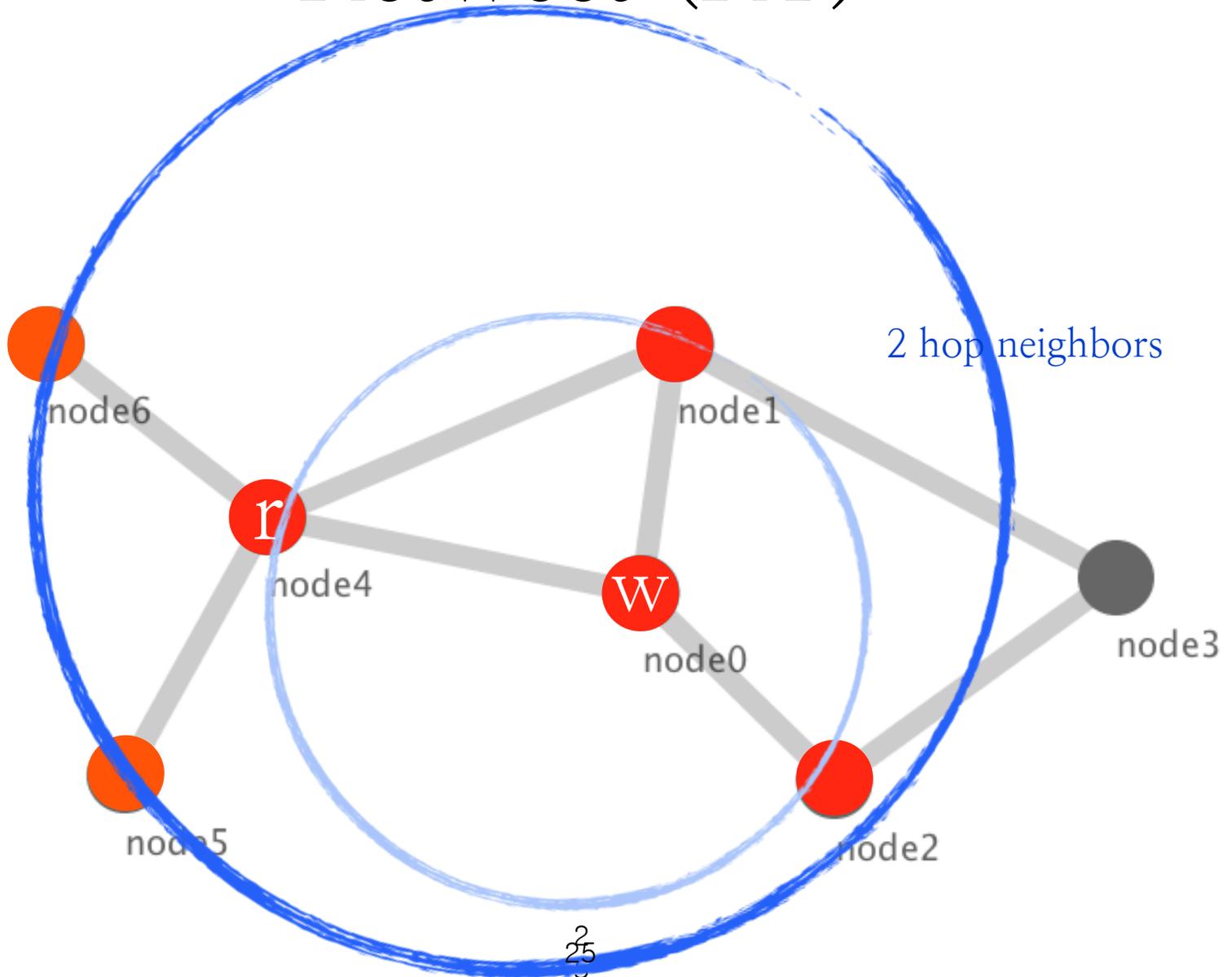


# Retweet (RT)

- Not only 1 hop neighbors

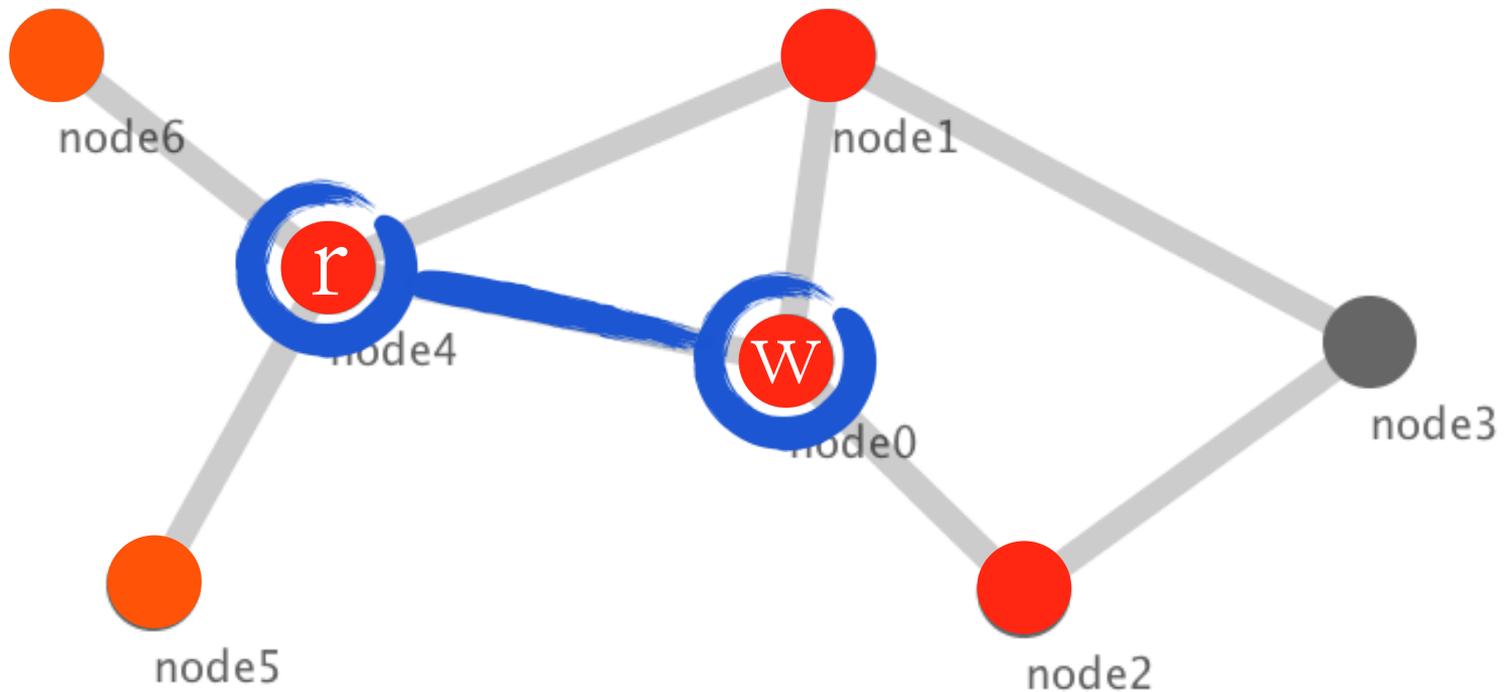


# Retweet (RT)

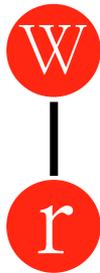


# We construct RT tree

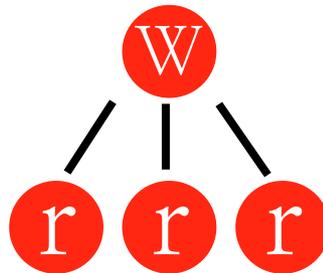
- A tree with writer and retweeter(s)



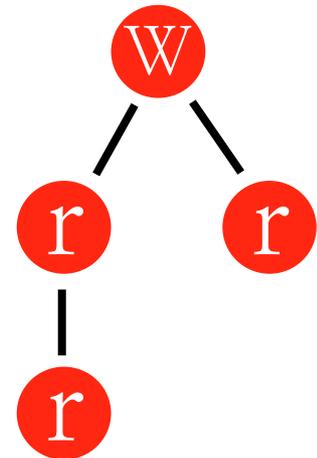
# Height of RT trees



1

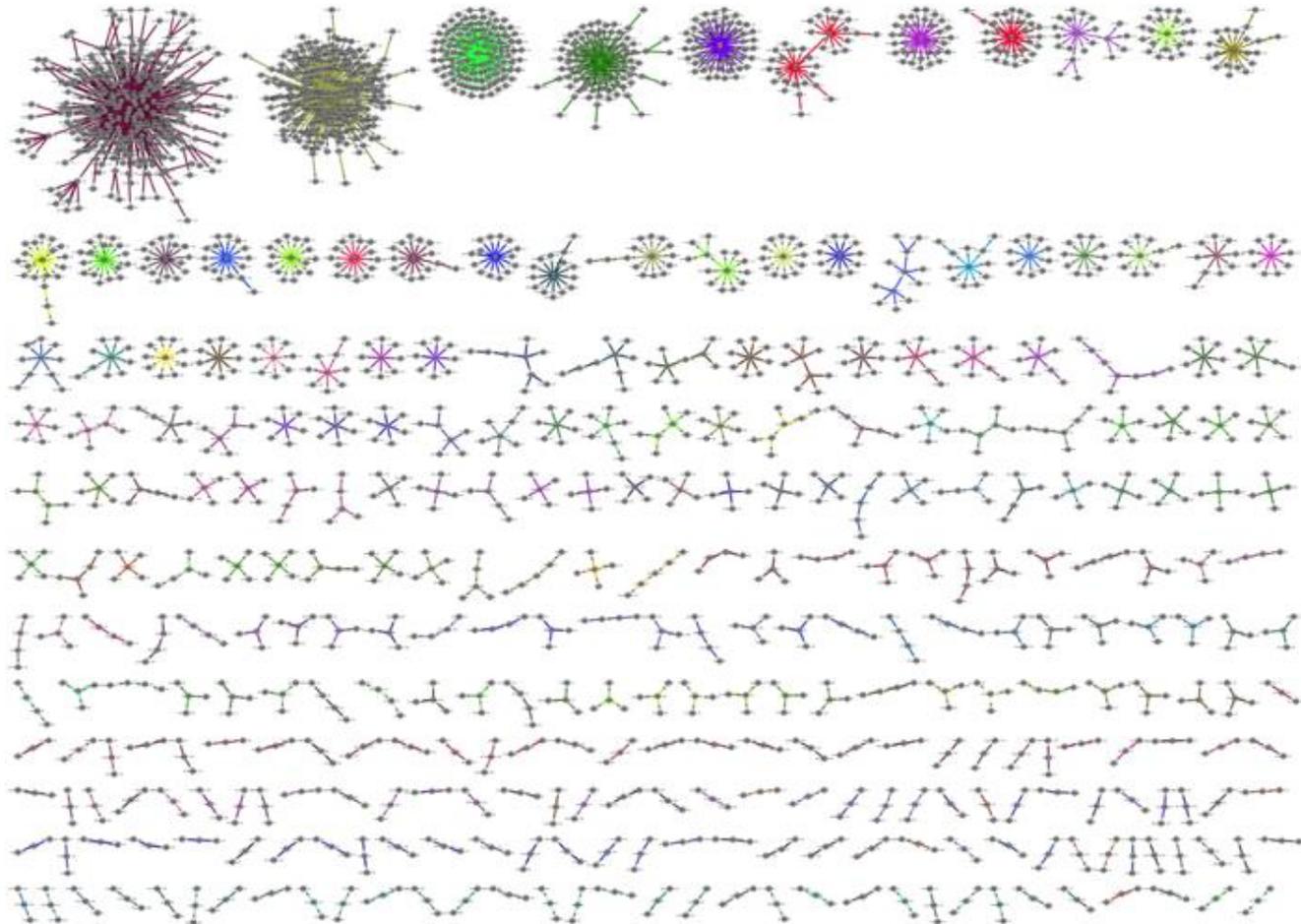


1

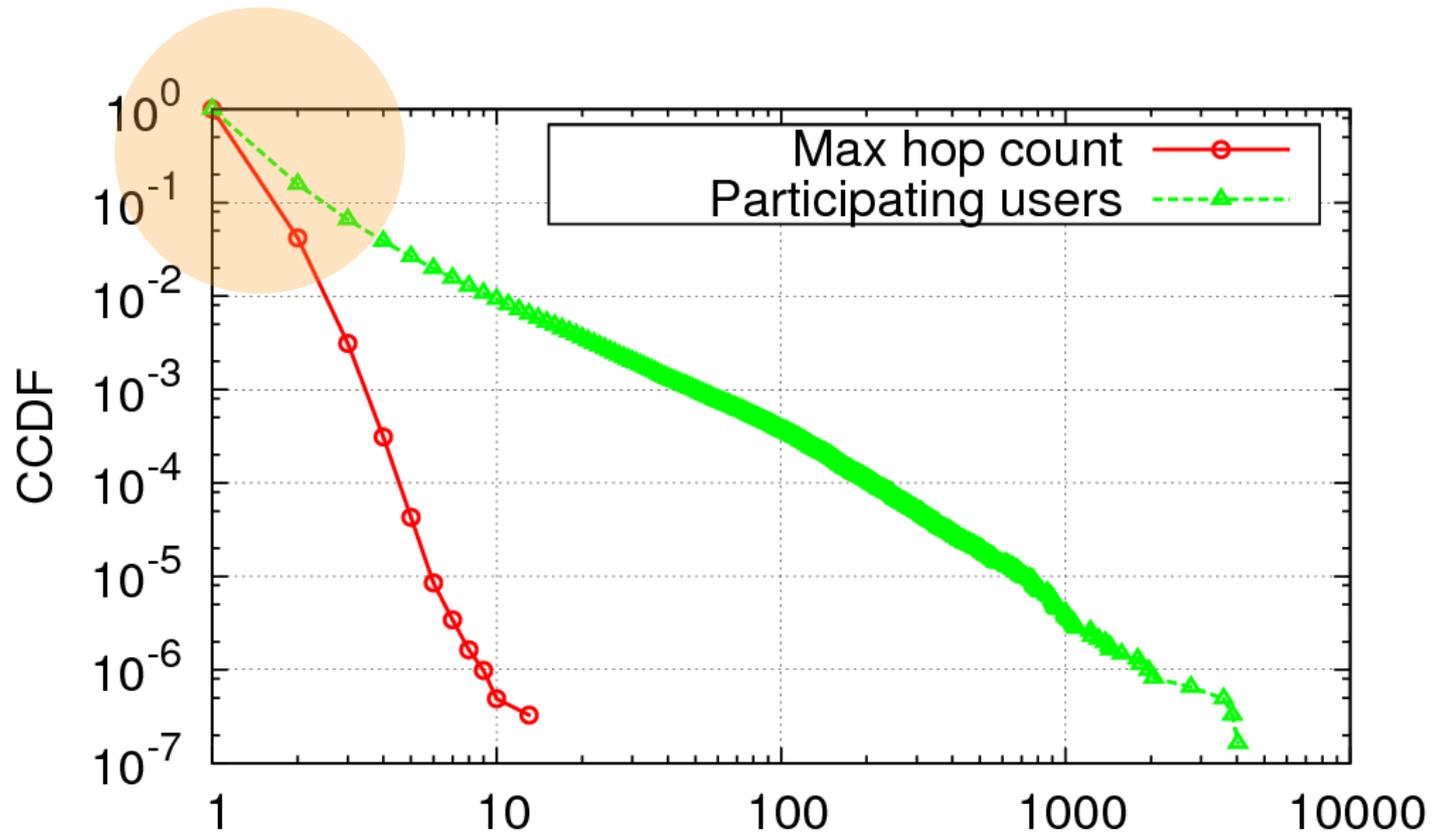


2

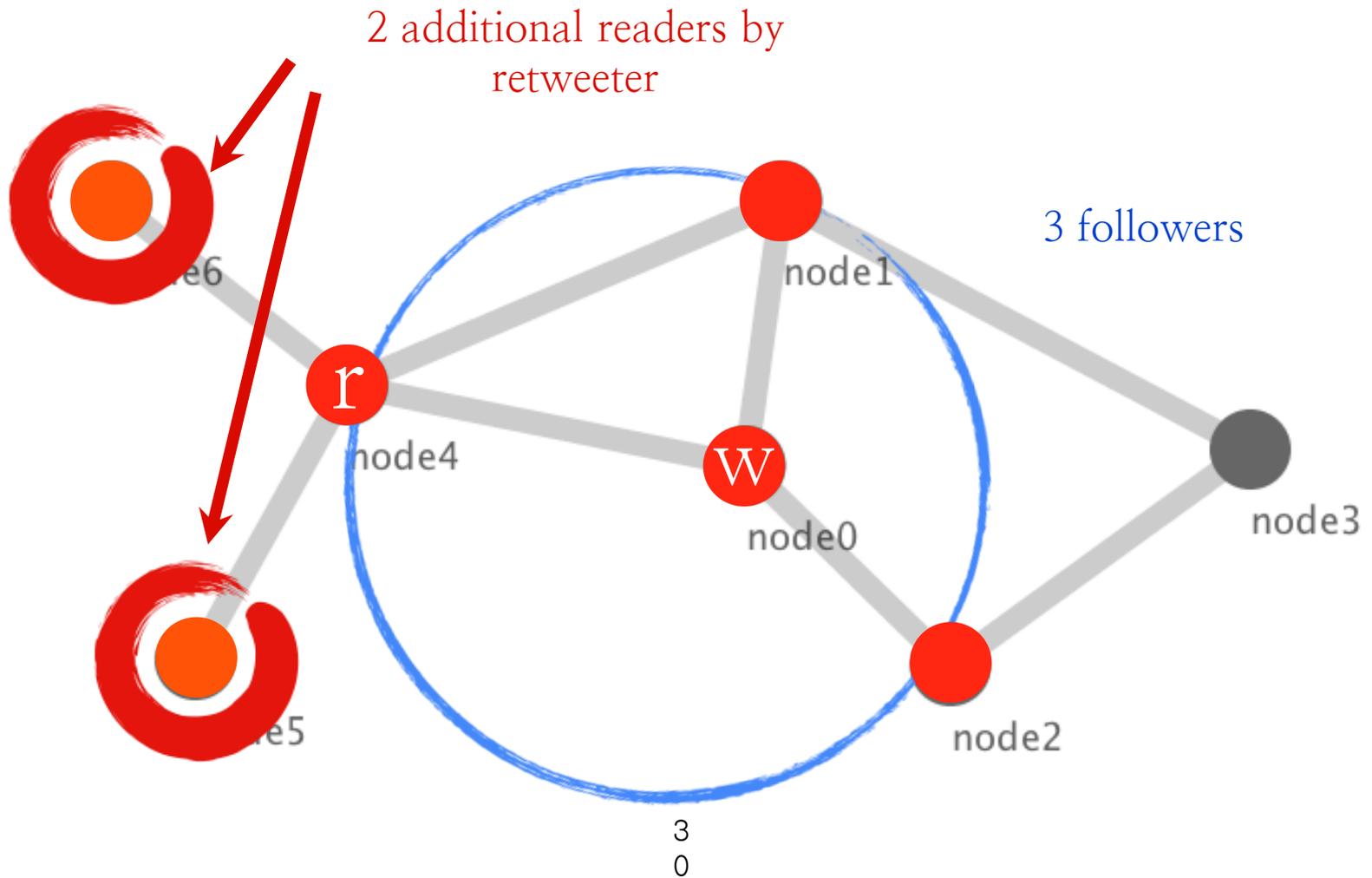
# Empirical RT trees



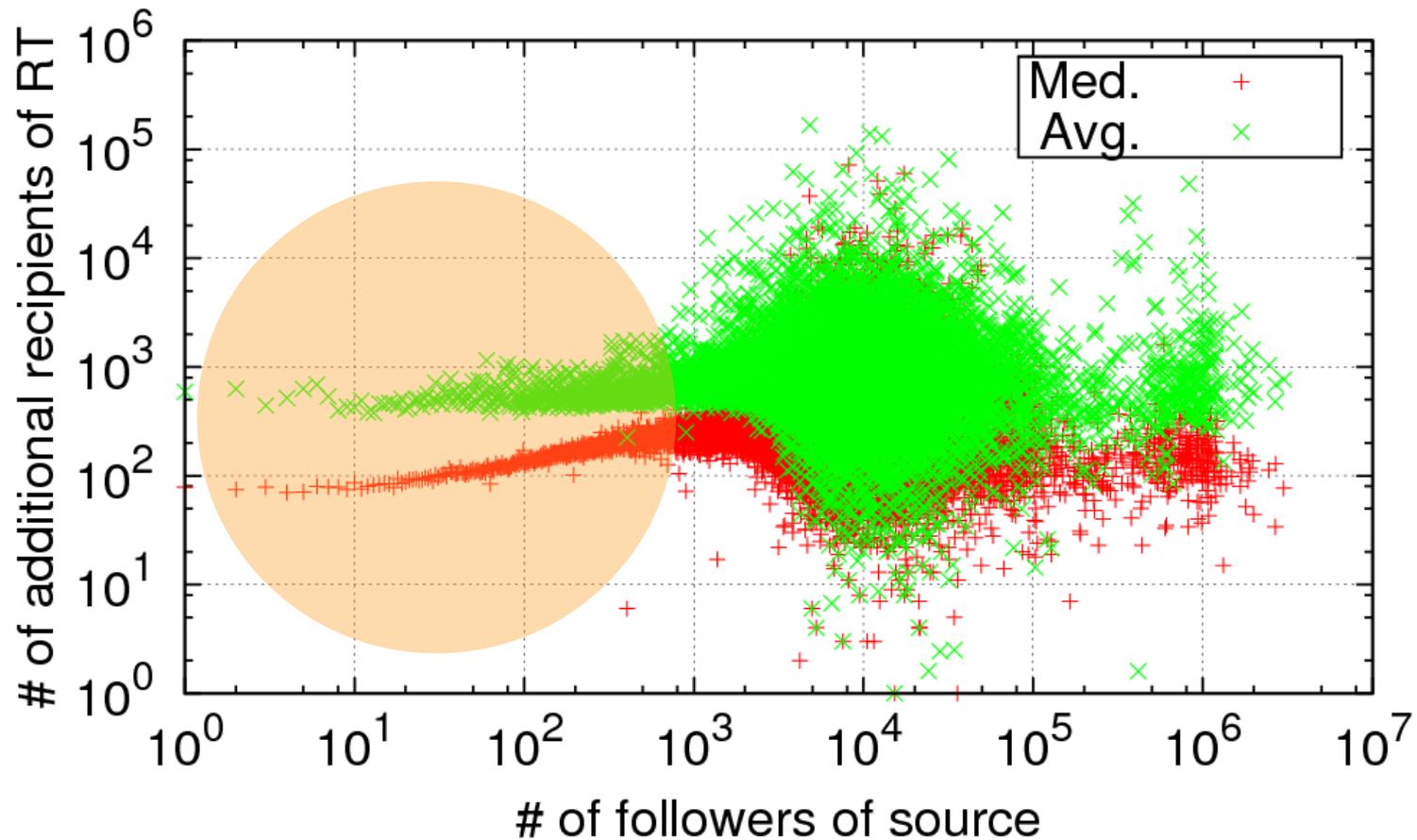
# 96% of RT trees = Height 1



# Additional readers



# 리트윗되면 정말 좋은가? 많은 사람들이 보나?



**그래서?????**

# 대규모 소셜 네트워크로 인한 변화

- facebook, twitter의 역할
  - 앱스토어와 같은 플랫폼으로 성장
  - 개인에 대한 엄청난 양의 정보 소유

- 자료의 절대량 비교

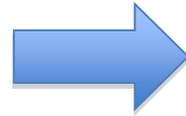
- Google vs Facebook

- 유튜브 영상

- vs 전화 통화량

$$10\text{kbps} \times 3600\text{sec} = 36\text{Mbits} = 4\text{MB}$$

# 의료 정보 대규모화

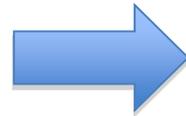


Physical observation



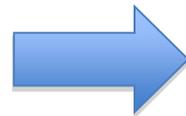
Blood pressure,  
Cholesterol,  
Medical images, etc

Phenotype



유전자 정보

Susceptibility,  
Static



네트워크바이오정보:  
유전체  
후성유전체  
단백질  
글라이켄

Susceptibility,  
Dynamics



# 대규모 소셜 네트워크로 인한 변화

- facebook, twitter의 역할
  - 앱스토어와 같은 플랫폼으로 성장
  - 개인에 대한 엄청난 양의 정보 소유
- 자료의 절대량 비교
  - Google vs Facebook
  - 유튜브 영상
  - vs 전화 통화량
  - $10\text{kbps} \times 3600\text{sec} = 36\text{Mbits} = 4\text{MB}$
  - vs 의료 자료
  - 1인당 100GB-10TB

# 유튜브 vs 의료 자료 vs Facebook/Twitter

- 차이점
  - 자료 공유 방법
  - 자료 공유 빈도
  - 자료 활용의 연산 복잡도
  - 자료 활용의 연산 빈도

**이쯤에서 미래 인터넷?**

# 현재 인터넷이 어때서??

- 한계점?
  - IPv4 주소 고갈, IPv6 널리 쓰이지 못하고 있다
  - DDoS 공격을 원천적으로 봉쇄할 수 없다
  - 이동통신성능은 좋지 않다

# PC로 만드는 “고성능” 장비



Backbone Core Router:  
100Tbps  
with ASIC and network processors



PC-based Software Router:  
10 to 40 Gbps  
with graphics cards

# 범용 PC 기반 고성능 장비 필요한가?

- 10Gbps 이더넷
  - 64B 패킷이라면 초당 2500만개
  - CPU 3GHz => 패킷당 100 cycle 겨우 할당

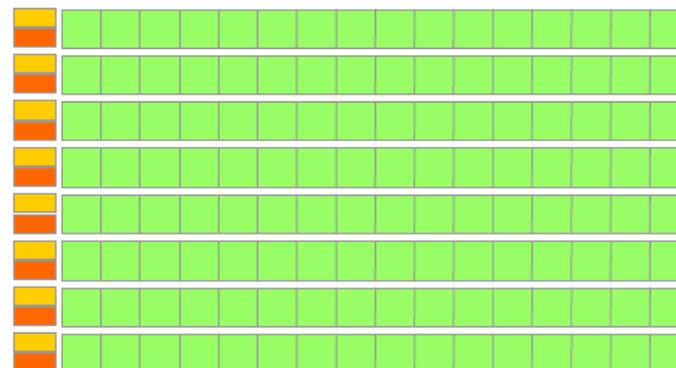
# CPU와 GPU의 구조적 차이



CPU: small # of super-fast cores



GPU: large # of small cores



**감사합니다**