

Twitter: Network properties analysis

1) Sobre qué va este paper

- Estudio de la topología de twitter
- Representación de grafo
- Mediciones sobre twitter
- Extrapolación de la estructura de twitter a partir de la muestra

Glosario:

- **SNS** (Social networking sites): comunidades virtuales donde se comparte información
- **web 2.0**: forma de acceso a la información a través de estas comunidades

2) Intro

La web "normal" se organiza en torno al contenido. La web 2.0 (SNS) se organiza en torno a los usuarios

¿por qué? son los usuarios los que publican su perfil y los generadores de contenido, pero sobre todo, crean links a otros usuarios con los que se asocian.

¿Qué podemos sacar de esto?

Podemos encontrar usuarios con intereses similares y encontrar contenido (conocimiento) que ya ha sido contrastado por otros usuarios.

SNS nos da una oportunidad para medir cómo las tendencias, las ideas y la información viaja a través de distintas comunidades sociales. Es decir, analizando la estructura en grafo podemos 1. Medir directamente el crecimiento de internet. 2. detectar usuarios "trusteds" 3. detectar usuarios influyentes

3) Base

Topología

- Los usuarios son nodos y establecen una relación (dentro de la red social) con otros usuarios (arista) por medio de los links.
De momento, no clasifica la relación y declara que puede ser de cualquier tipo (amistad, conocido, interés similar)
- También tenemos los grupos de usuarios que comparten algún interés

Network dynamics

1. **Dynamics of the network** se refiere al cambio de la estructura de la red en sí.
Por ejemplo añadir nuevos nodos (usuarios) o añadir o quitar relaciones (ties) entre ellos; en este caso medir la red se refiere a estudiar la red en el tiempo para ver su evolución (este es el tipo que le interesa al autor)
2. **Dynamics on the network** Los nodos (usuarios) interactúan entre ellos y esto está condicionado por lo que hacen sus vecinos. No le interesa al autor.

Me parecen dos términos muy interesantes

El dato

En un año (marzo 2008 - marzo 2009) twitter creció un 2565% (hablando en términos de usuarios)

Analizando Twitter

Es muy fácil mapear los usuarios como nodos y los links directos como relaciones. Esto nos da un grafo.

Nota: la información fluye en sentido contrario al link en sí.

$A \Rightarrow B$

La información que publica B, es escuchada por A

Cómo medir redes sociales

La técnica se basa en cuán importante es la relación entre dos nodos.

1. Grafos (o sub-grafos) **Strongly connected** Desde cualquier nodo existe un camino a cualquier otro nodo). Vamos a dividir el grafo en sub grafos que sean fuertemente conectados
2. **in-degree** número de aristas entrantes: mide la autoridad de un nodo con respecto al tiempo (a más followers, más influyente)
3. **out-degree** nos va a servir para medir el coeficiente de clustering
Nos dice cuán cerca está un vertice y sus vecinos de convertirse en un grafo completo.
Si el coeficiente de clustering aumenta, indica una tendencia de un grafo a convertirse en interconectado. (nodo A directamente conectado con todo otro nodo)
4. **Densidad de la red** número de aristas - relaciones / todas las relaciones posibles; si nos acercamos a 1.0 está muy densa la cosa
5. **Betweenness** Un nodo A tiene más betweenness que otro nodo B si tiene caminos cortos con otro vértice
6. **Closeness** A mayor closeness, un nodo recibirá la información más rápidamente
7. **Following ratio** followers / following
 - $\sim 0 \implies$ web-bot or spider coleccionando información sobre trending topics
 - $< 1 \implies$ buscamos coleccionar información
 - $1 \implies$ standard
 - $1 \implies$ generador de contenido apreciado por su propia comunidad
 - $10+ \implies$ nodos jefes Huge impact around general media

4) El trabajo

Se hace un grafo de twitter, escoge un nodo super poderosos con un rating de 30 y empieza a dibujarlo.

Toma 14.148 (qué número tan raro) nodos

La información que recoge de cada usuario

nick	messages	out-degree	in-degree

A partir de estos datos, queremos sacar

- following ratio
- betweenness
- clustering coefficient
- closeness
- sub-grafos strongly connected
- densidad e la red

Todos los datos los va a sacar de 4 capturas (fotos) en el tiempo

me llama mucho la atención que esas capturas las hace con muy poco periodo entre ellas: 11 abril 22 abril 23 abril 29 abril

5) Conclusiones a las que llega

1. 80% unos 1500 mensajes; 5% más de 5000 mensajes (los Danis)
2. Los usuarios activos postean de forma regular
3. La media es de 9 posts diarios / user
4. 25.5% tiene 50 followers
5. al llegar a la cifra de 600 followers, la cifra de followers se dispara a los 100.000

Sobre el out-degree

1. un usuario nuevo sigue mucho y luego va refinando
2. Los bots aparecen con las tendencias
3. hay follow back (joder y para eso un estudio)

Rating

1. el 50% tiene un 1:1
2. No hay apenas cambios en este ratio. Y solo se aprecian cambios con los usuarios a los que sigue mucha gente (supongo que por el disparo

... usuarios a los que sigue mucha gente (respaldo que por el momento del que habló al principio)

3. El rating lo usa para identificar lo que llama las hojas del grafo (usuarios inactivos de twitter) con cuenta pero sin uso ==> 45%!!!!

Me parece muchísimo

Sobre clustering

1. Los usuarios con mucho clustering son los que se enteraran de los eventos más rápido que nadie.
2. Saca en conclusión que el 30% de los usuarios tiene buen clustering (recibirá la información directamente o con un retweet máximo)

Betweenness

1. 20% de los usuarios tienen este coeficiente nulo a casi nulo, es decir, no son importantes en la red. (pero si dijo que las hojas eran como el 40%?!)

This descriptive measurement shows that 20% users have very low or inexistent betweenness coefficient, meaning that those users do not have a relevant role within the network. In other words, they have a very low in-degree and out-degree that might show the real amount of inactive users.

On the other hand, high betweenness means that the following ratio tends to be 1; users follow and are being followed. It should be noticed that betweenness differs from clustering coefficient; the first one shows 50% of the users "playing a role" within the network, while the second one shows that 30% of the total nodes is clusterable or tend to be in communities.

Closeness

1. Esta medida está entre 0.22 y 0.3 entre usuarios muy conectados.
2. La usa para identificar a los usuarios que se acaban de dar de alta, que son el 8%

Densidad de la red

1. Que los grupitos no se hacen de golpe y se van formando y modificando.
2. Un nodo - a menos que sea nuevo - pertenece a varios strongly group.

(Otras conclusiones a las que no habría llegado)