



Ciencia de Datos y BigData

Análisis de Redes Sociales

Dr. José Ramón Iglesias

DSP-ASIC BUILDER GROUP
Director Semillero TRIAC
Ingeniería Electronica
Universidad Popular del Cesar

Basado en filminas de Paola Velardi

<http://twiki.di.uniroma1.it/twiki/view/Estrinfo>

¿Qué podemos minar en redes sociales?

Contenido:

- ¿De qué se está hablando? → análisis y visualización
- ¿En qué temas podemos agruparlo? → clustering
- ¿Con qué opiniones, con qué sentimientos? → clasificación

Estructura:

- ¿Qué grupos hay?
- ¿Cómo se propaga la información?
- ¿Quiénes tienen más influencia?
- ¿Quiénes tienen más prestigio?

¿Qué podemos minar en redes sociales?

Contenido:

- ¿De qué se está hablando? → análisis y visualización
- ¿En qué temas podemos agruparlo? → clustering
- ¿Con qué opiniones, con qué sentimientos? → clasificación

Estructura:

- ¿Qué grupos hay?
- ¿Cómo se propaga la información?
- ¿Quiénes tienen más influencia?
- ¿Quiénes tienen más prestigio?

Medidas sobre nodos

Medidas sobre nodos

- Alcance
- Buzz
- Influencia
- Sentimiento

Medidas sobre nodos

- Alcance
- Buzz
- Influencia
- Sentimiento

A cuántas personas llega el mensaje

- Visitas a una página, post
- Likes
- Republicaciones, menciones
- Seguidores
- Engagement: una métrica que acumula likes, seguidores, menciones... a gusto

Medidas sobre nodos

- Alcance
- **Buzz**
- Influencia
- Sentimiento

Amplificación de un mensaje a través de los media: cuánto hablan de eso, cuánto lo mencionan, dónde lo mencionan, cómo lo mencionan y por qué

Medidas sobre nodos

- Alcance
- Buzz
- **Influencia**
- Sentimiento

Probabilidad de que otros usuarios vean, faveen, republiquen, comenten

- Ampliación: Retwittes (twitter), comparticiones (facebook, youtube)
- Aplauso: likes

Medidas sobre nodos

- Alcance
- Buzz
- Influencia
- **Sentimiento**

Con qué sentimiento se está conversando

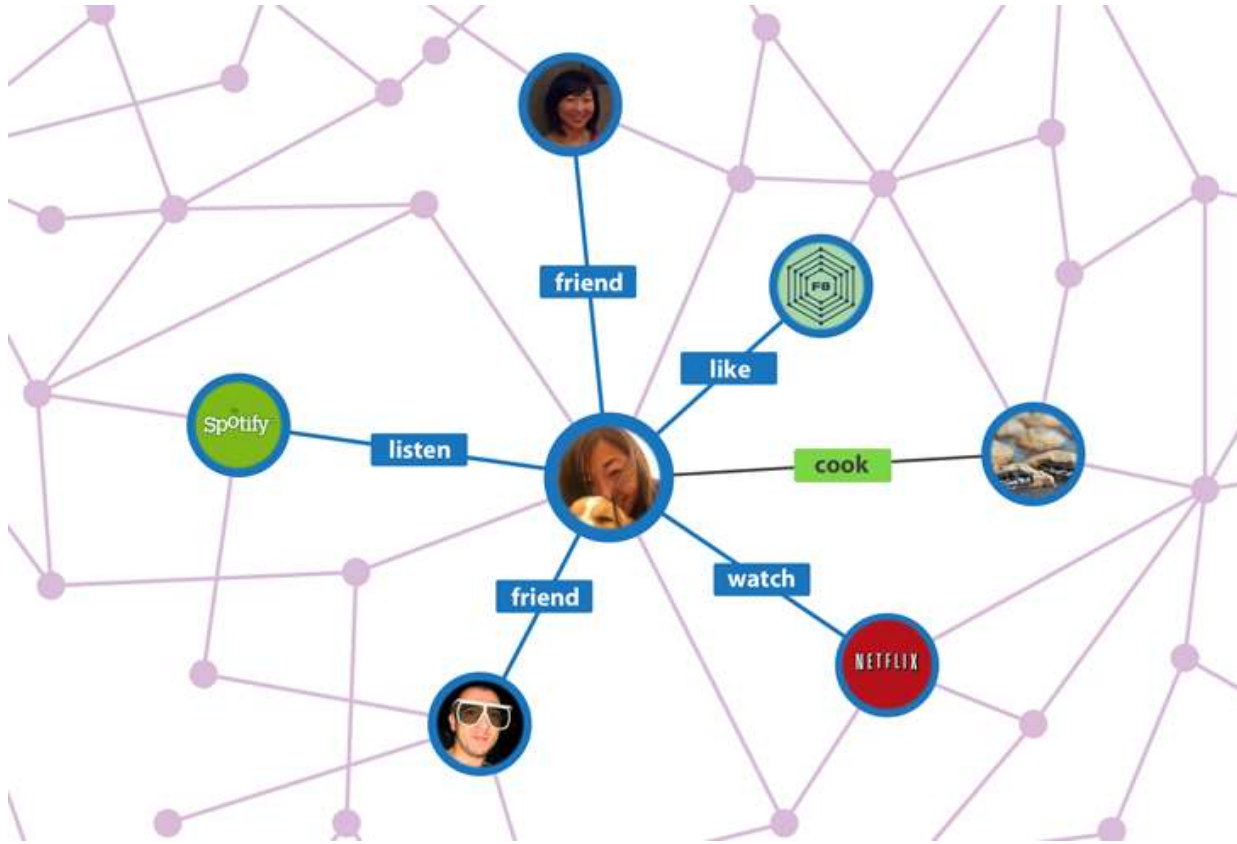
- Sentimiento: positivo / negativo
- Emoción: enojo / alegría /...
- Intensidad
- Clases específicas

Medidas sobre grafos

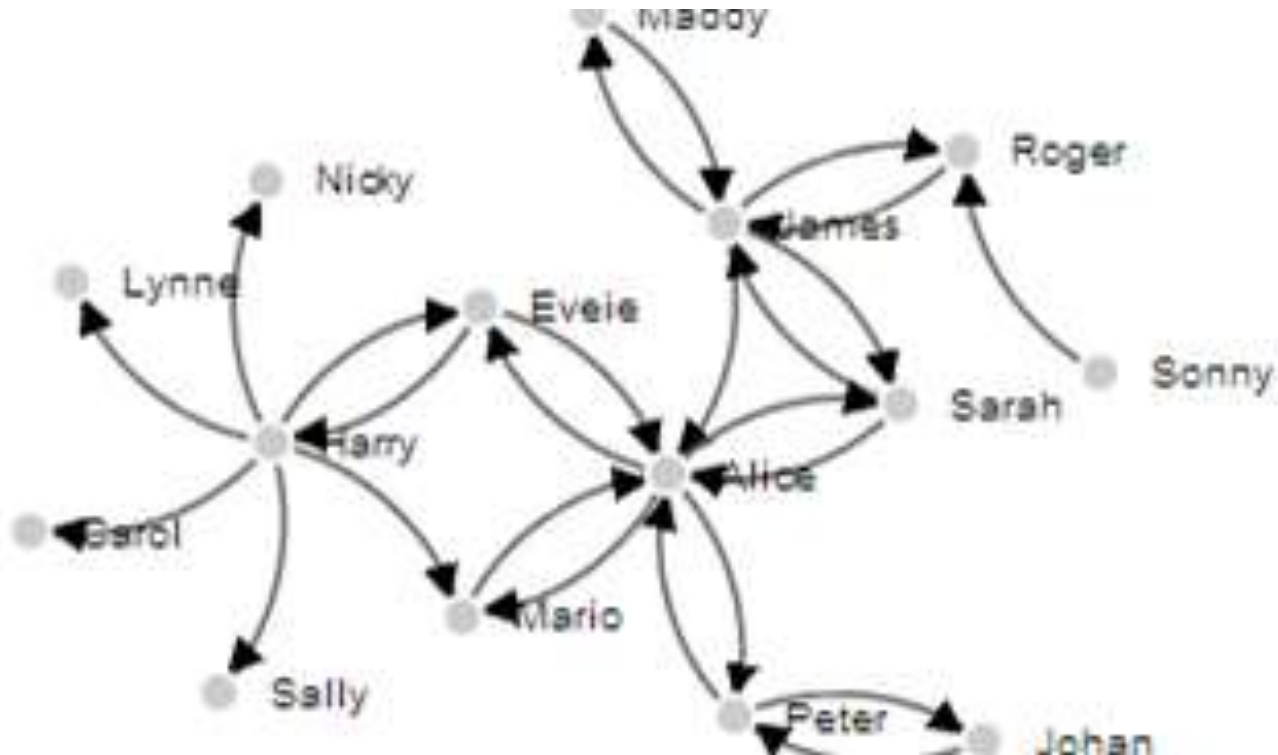
Capacidades de medidas sobre grafos

- Más profundas (menos superficiales) que las de nodos
- Detectar comunidades
- Detectar flujo de información
- Detectar actores influyentes

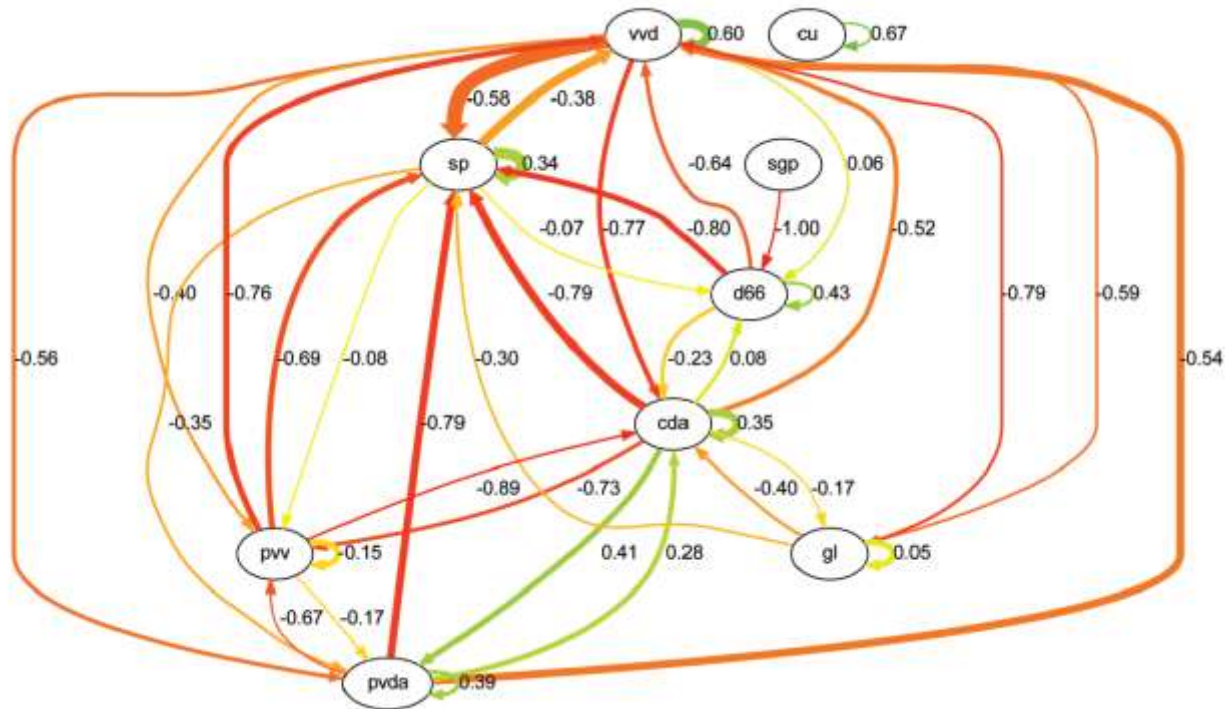
Redes sociales como grafos



Redes sociales como grafos: twitter



Redes sociales como grafos: completo

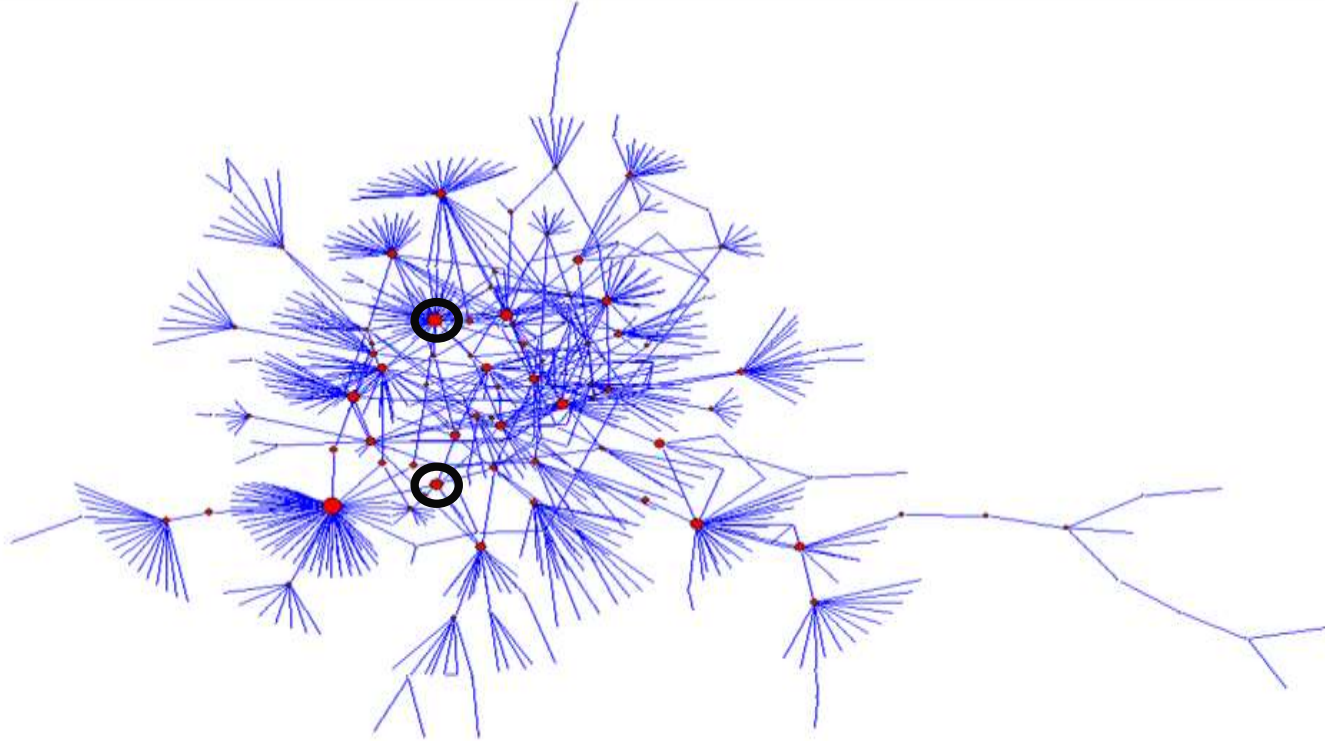


Jugadores clave

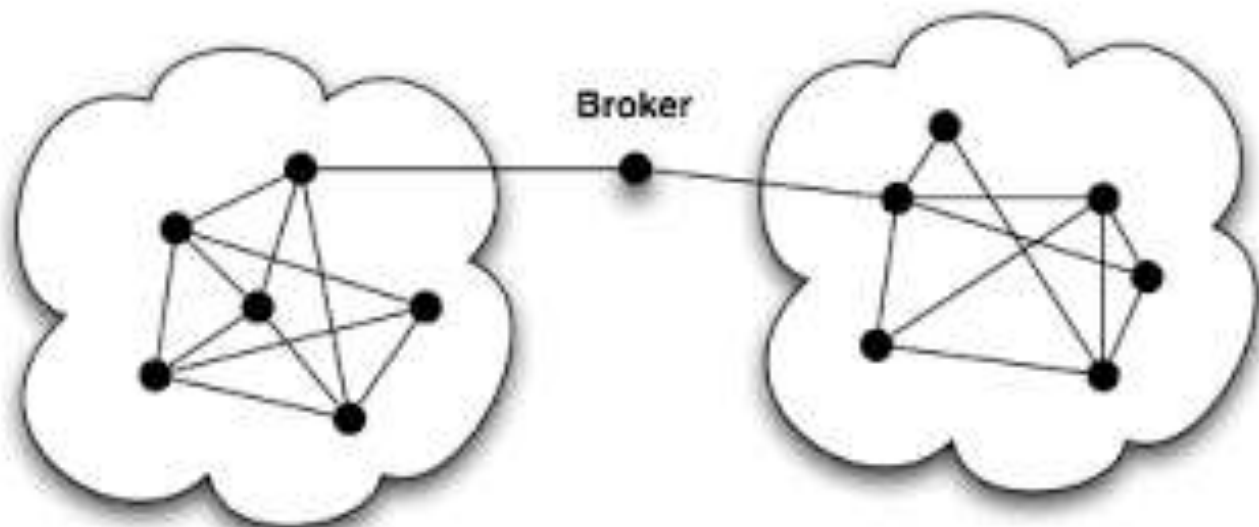
Basado en conectividad para encontrar

- **Centralidad** (no dirigido)
 - Identificar los nodos que están en el centro (vs. periferia)
 - Grado (*degree*): nodos con mayor número de arcos
 - Cercanía (*closeness*): inversa de la distancia al resto de nodos
 - Camino (*betweenness*): el nodo se encuentra en el camino más corto entre otros dos nodos
- **Prestigio** (dirigido)
 - Influencia (arcos salientes) (Pagerank, Hubs and Authorities)
 - Soporte (arcos entrantes)
 - Brokers (puentes de conectividad en el grafo, si se sacan, reducen la conectividad del grafo)

Betweenness centrality



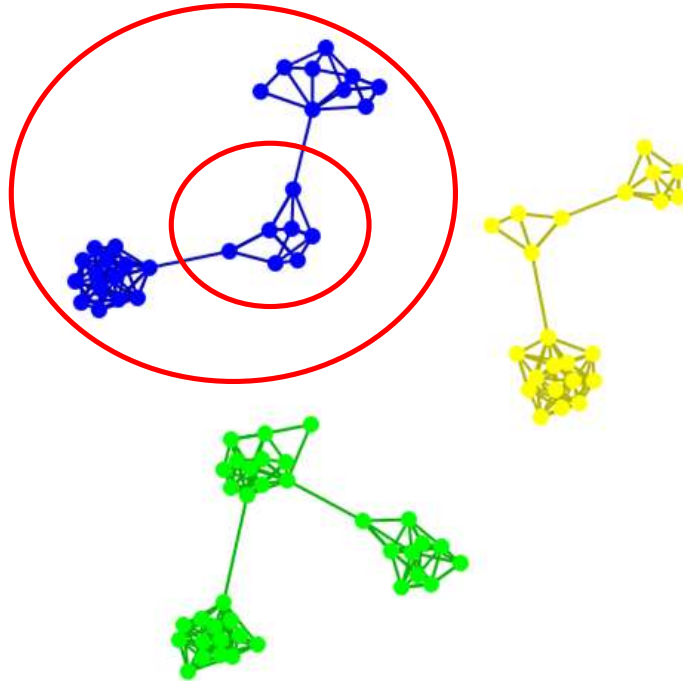
Brokers



Medidas sobre grafos:
Detección de comunidades

Detección de comunidades

Grupos de nodos que interactúan entre ellos más que con el resto

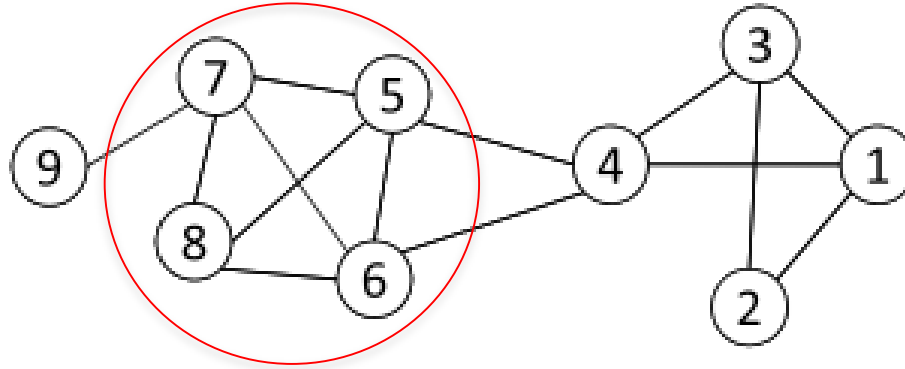


Métodos para detección de comunidades

- Nodo-céntricas: todos los nodos satisfacen ciertas propiedades
- Grupo-céntricas: se consideran las conexiones dentro de un grupo como un todo, el grupo tiene que satisfacer ciertas propiedades
- Grafo-céntricas: se parte el grafo en partes disjuntas
- Jerárquica: se subdivide el grafo iterativamente

Comunidades nodo-céntricas

Encontrar cliques: subgrafos donde todos los nodos son adyacentes entre ellos

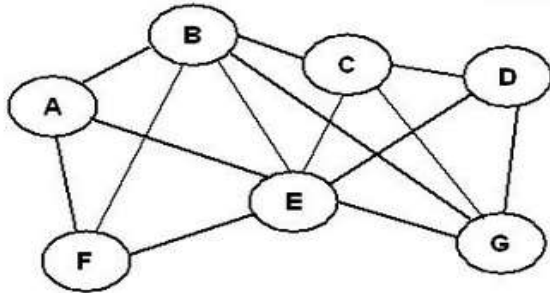


Es costoso, pero se puede hacer aproximado

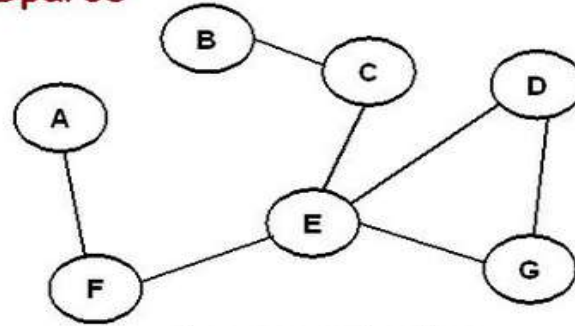
Comunidades grupo-céntricas

Todos los nodos del grupo satisfacen la propiedad de mantener un cierto nivel de conexión entre ellos

Dense vs. Sparse



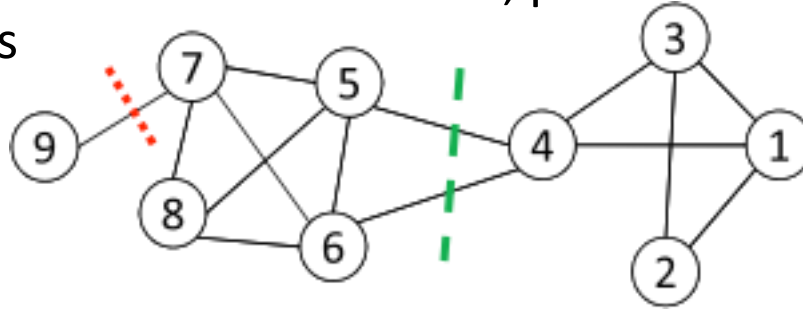
$$\frac{2E}{V(V-1)} = \frac{28}{49}$$



$$\frac{2E}{V(V-1)} = \frac{7}{49}$$

Comunidades grafo-céntricas

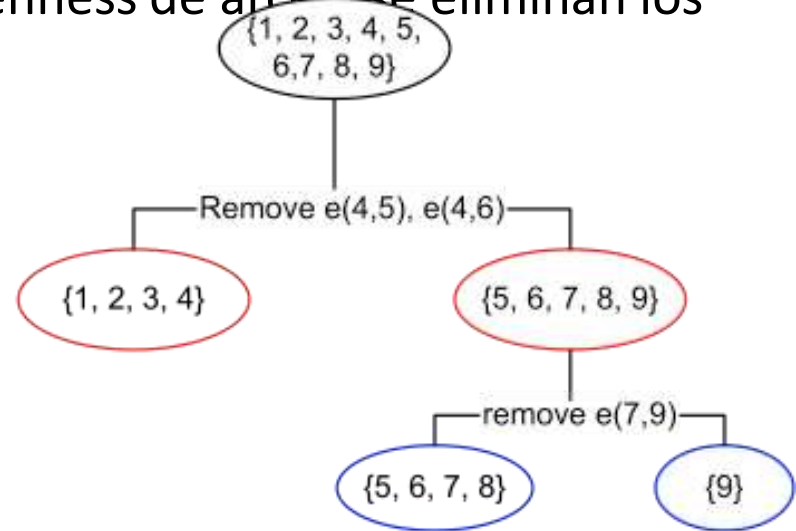
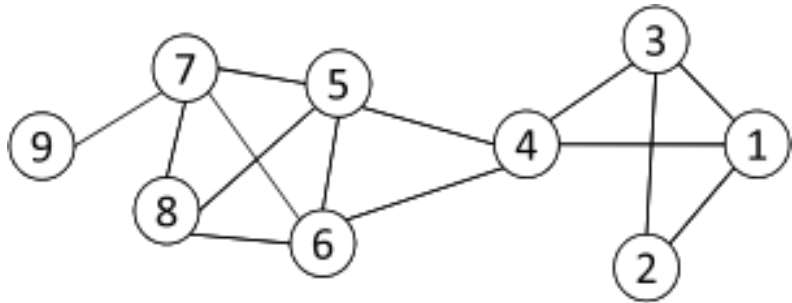
- Clustering basado en semejanza entre nodos caracterizados por su vecindad
- Clustering basado en mincut: se cortan la menor cantidad de arcos posibles, por lo tanto, nos quedamos con clusters con muchos arcos entre ellos y pocos arcos hacia afuera, pero suele dar cortes desbalanceados



- Mejor: clustering basado en Radio Cut o Normalized Cut

Comunidades grafo-céntricas

- Clustering basado en semejanza entre nodos caracterizados por su vecindad
- Clustering basado en mincut, Radio Cut o Normalized Cut
- Clustering divisivo basado en betweenness de arcos: se eliminan los arcos con más betweenness



Comunidades grafo-céntricas

- Clustering basado en semejanza entre nodos caracterizados por su vecindad
 - Clustering basado en mincut, Radio Cut o Normalized Cut
 - Clustering divisivo basado en betweenness de arcos: se eliminan los arcos con más betweenness
-
- Maximización de modularidad: busca la división en grupos que maximiza una medida grupo-céntrica (p.ej., conectividad), como el Louvain Modularity (el que implementa gephi)

Algunas implementaciones

<http://socialmediadata.org/social-media-research-toolkit/>

<https://gephi.org/>

<https://github.com/eflegara/NetStruc/blob/master/6.%20Community%20Detection.ipynb>