

Aplicaciones y Casos Prácticos de Redes

Inteligencia Colectiva y Sistemas de Recomendación

Master Propio en Data Science y Big Data

Agenda

- Muestreo de Redes
 - Estrategias de Muestreo
 - Percolación y Muestreo
 - Infección y Muestreo
- PageRank aplicado a Sistemas de Recomendación
- Análisis de Redes Sociales
- Algoritmos de representación
- Software

Muestreos Aleatorios en Redes

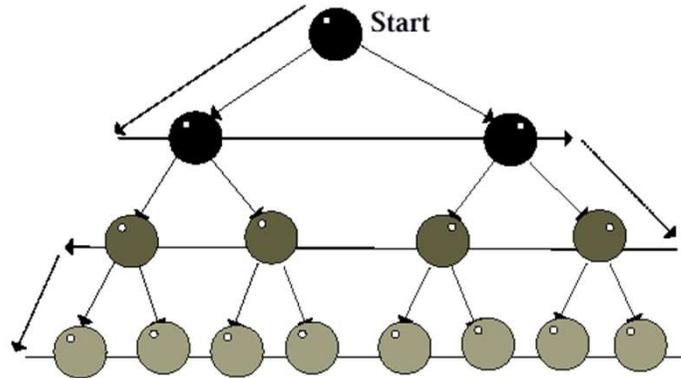


- Escenarios:
 - Tenemos una red GRANDE que cabe en memoria, pero queremos ejecutar un algoritmo sobre ella que puede necesitar muchos recursos... ¿cómo podemos acelerar la computación?
 - Tenemos una red GRANDE que cabe en disco, pero no en memoria... ¿cómo podemos analizarla en un tiempo razonable?
 - Es muy costoso, o imposible, recolectar todos los datos de una red, y solo podemos acceder a ella por medio de crawling... ¿podemos inferir resultados de la red completa analizando partes de la red?
- En todos estos escenarios necesitamos hacer un **muestreo** de la red.

Muestreos Aleatorios en Redes



- Estrategias de muestreo:
 - Selección aleatoria de nodos (únicamente si se tiene acceso a toda la red)
 - Selección aleatoria de aristas (únicamente si se tiene acceso a toda la red)
 - Basado en crawling:
 - BFS, DFS, ...
 - Caminos aleatorios



- En todos estos escenarios necesitamos hacer un **muestreo** de la red.

Muestreos Aleatorios en Redes



- Objetivos:
 - Obtener una subred representativa (scale-down)
 - Estimación de un parámetro de la red
 - Estimación de atributos de los nodos
 - Estimación de atributos de las aristas
- Algunas estrategias de muestreo funcionan mejor que otras para algunos casos.

Muestreos Aleatorios en Redes



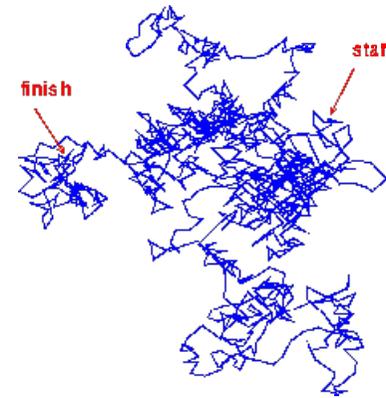
- Selección aleatoria de nodos:
 - Muestreo uniforme: no mantiene las redes libres de escala
 - Muestreo basado en el grado
 - Muestreo basado en el pagerank
- Selección aleatoria de aristas:
 - Muestreo uniforme
 - Muestreo uniforme de nodo-aristas
 - Híbrido de los anteriores
 - Muestreo inducido (uniforme + completar aristas de nodos presentes)



Muestreos Aleatorios en Redes

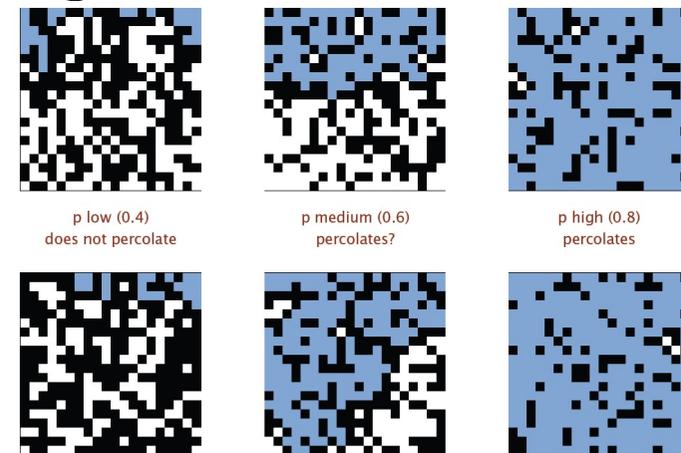


- Basado en Crawling:
 - BFS: tiende a sobremuestrear nodos de grado alto
 - DFS : tiende a sobremuestrear nodos de grado alto
 - Camino aleatorio : tiende a sobremuestrear nodos de grado alto
 - Modificación de la probabilidad de transición para evitar el sobremuestreo
 - FFS (versión probabilística de BFS)
 - Expansión
 - Camino aleatorio con saltos aleatorios
- Muestreos con crawling de redes aleatorias suelen ser libres de escala.



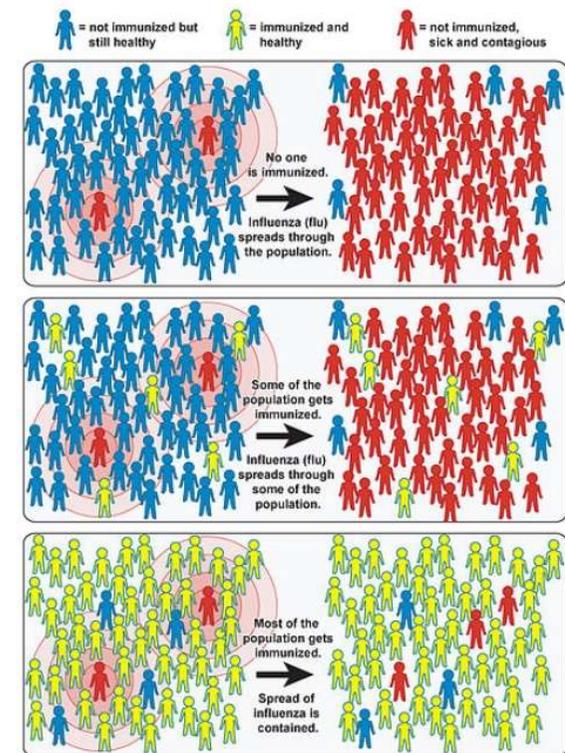
Percolación, Resiliencia y Muestreo

- Con probabilidad p vamos eliminando nodos de una red. El objetivo es ver si tras el proceso la red mantiene el tamaño de sus componentes conexas (**percolación**).
- La forma en que los nodos son seleccionados para ser eliminados puede ser determinante para la **resiliencia** de la red. Suele aparecer una probabilidad crítica para que la red pierda sus características:
 - Este valor cambia radicalmente si se hace un muestreo uniforme o no (ataques al azar vs. ataques dirigidos)



Difusión, Infección y Muestreo

- En los modelos de difusión/infección puede ser determinante el comportamiento asintótico de la dinámica según las características de los nodos inicialmente informados/infectados.
- Casos de uso:
 - Publicidad
 - Ataques
 - Rumorología
 - Opinión/Voto

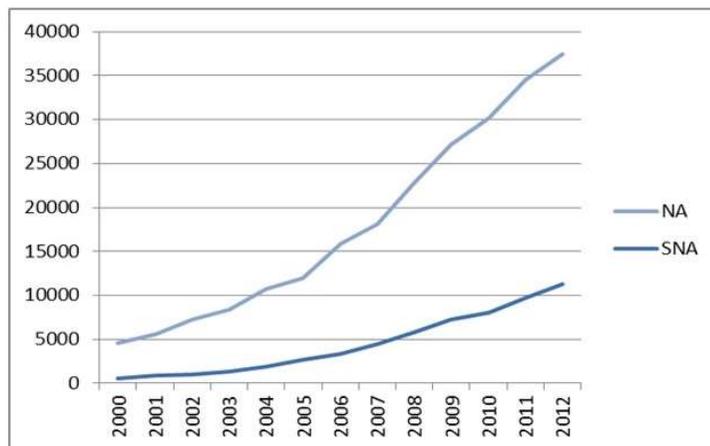


Análisis de Redes Sociales



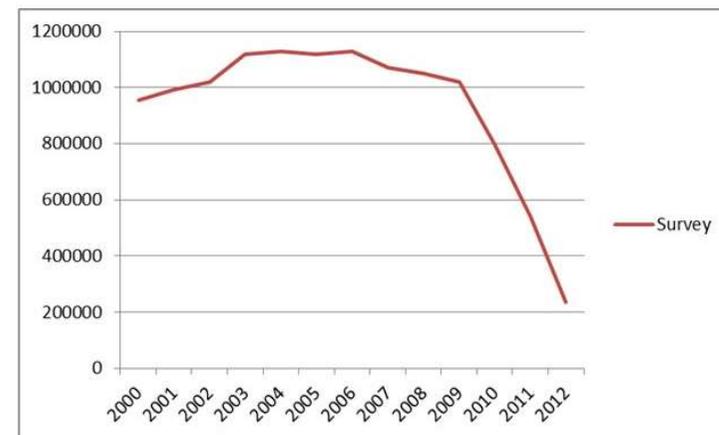
Aproximación metodológica y teórica que enfatiza el estudio de las relaciones entre actores, tanto relaciones entre personas, organizaciones, países o cosas, proyectando la información social en forma de redes.

- Uso de análisis de redes en investigación



Número de artículos o libros de investigación donde se nombre análisis de redes "networks analysis" (NA) o análisis de redes sociales "social network analysis (SNA)

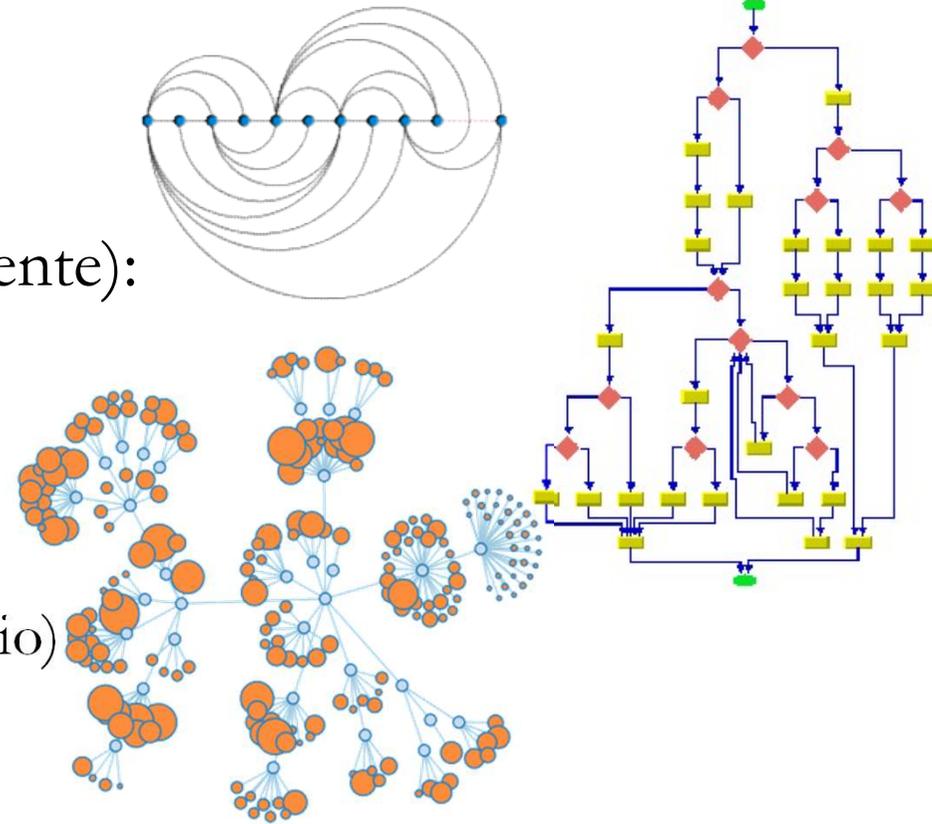
- Uso de método de encuesta en investigación



Número de artículos o libros de investigación donde se nombre encuesta "survey"

Algoritmos de Representación

- Representación geométrica de las redes (normalmente en 2D) con el fin de mostrar características internas de la red visualmente.
 - Distribución (Layout)
 - Estilos en nodos (tamaño, color, forma)
 - Estilos en aristas (grosor, color, patrón)
- Objetivos (no satisfactibles simultáneamente):
 - Minimizar cruces
 - Minimizar área
 - Minimizar solapamientos
 - Maximizar simetrías
 - Verificar restricciones (dependen del dominio)



Software



- NetworkX (Python)
- NetLogo
- iGraph (R, C++, Python)
- Network & Statnet (R)