Universidad de Chile
Facultad de Ciencias Físicas y Matemáticas
Profesores: José Correa, José Verschae
Prof. Auxiliar: Alberto Vera Azócar

Fundamentos de Redes Sociales
Clustering y Modelos de Redes

18 de diciembre de 2013

Problema 1 [Modelo Estructural].- Considere el siguiente modelo de grafo aleatorio: en el instante 0 existe un nodo v_0, en el instante t se agrega un nodo v_t y se agrega un arco desde v_t hacia cada nodo anterior de manera independiente y con probabilidad $p_t = p(1-e^{-t})$ donde $p \in (0,1)$ es un parámetro del modelo.

1. Dé una expresión para el grado esperado de un nodo al azar en tiempo t y muestre que es $\Omega(t)$
2. Estudie si el grado de los nodos sigue una ley de potencia.
3. Vea que para t grande se tiene una componente gigante conexa, es decir, con alta probabilidad existe una componente conexa de tamaño $\Omega(t^\delta)$
4. Estudie para que valores de t y α el nodo v_t pertenece a una α-comunidad fuerte.

Problema 2 [Clustering de Correlación].- En una red social podemos observar si las personas tienen una característica común (son amigos, les gusta la misma música, etc) o bien si tienen características disímiles.
Sea $G = (V, E)$ un grafo completo con etiquetas $+$ ó $-$ en los arcos. Recordemos que el agreement de una partición está dado como el número de $+$ dentro de los cluster más el número de $-$ que los cruzan.
Queremos encontrar un algoritmo que aproxime el problema de max agreement, denotemos OPT la partición óptima y $|OPT|$ el número de agreements. Finalmente para $V_1, V_2 \subseteq V$ denotamos $\delta^+(V_1, V_2)$ como el número de arcos etiquetados con $+$ que van de V_1 a V_2.

1. Sea $\varepsilon > 0$ dado, $OPT(\varepsilon)$ la partición óptima bajo la restricción que cada cluster no síngleton sea de tamaño mayor a εn, muestre que $|OPT(\varepsilon)| \geq |OPT| - \varepsilon n^2/2$
2. Sean $C_1, \cdots, C_k$ los cluster no síngleton de $OPT(\varepsilon)$ y C_{k+1} la unión de todos los síngleton. Denotemos $s_i = |C_i|$ y $e_{ij} = \delta^+(C_i, C_j)$. Explique por que
$$|OPT(\varepsilon)| = \left(\sum_{i=1}^{k} e_{ii}\right) + \left(\binom{s_{k+1}}{2} - e_{k+1,k+1}\right) + \left(\sum_{i\neq j}(s_i s_j - e_{ij})\right)$$
3. Existe un algoritmo $\mathcal{A}$ tal que, dados valores s_i y e_{ij}, $\mathcal{A}(s_i, e_{ij})$ retorna una aproximación de la mejor partición que satisfaga aquellas restricciones (no pruebe esto, use $\mathcal{A}$ como oráculo).

   Si la aproximación de $\mathcal{A}$ es tal que $OPT_{\mathcal{A}(s_i,e_{ij})} \geq OPT_{s_i,e_{ij}} - \varepsilon n^2/2$, proponga un algoritmo $\mathcal{B}$ para el problema de clustering tal que $OPT_{\mathcal{B}(G)} \geq OPT_G - \varepsilon n^2$.

Nota: Discutamos de como se puede usar el resultado anterior en un grafo G' no completo de la vida real.