

Trabajo 2 - Redes Sociales :

Comparación con modelos nulos

Juan Illanes Vásquez^{*1}

¹CICS, Facultad de Gobierno, Universidad del Desarrollo

1 Estadísticas descriptivas

La red que se va a trabajar será la pseudo proyección de la red bipartita comuna–empresa, según la metodología HH. Esta es una red con pesos, cuyos enlaces se interpretan como la mínima probabilidad condicional de que las empresas p y q hayan ganado una licitación a lo largo del territorio nacional en el mes de enero de 2025. Cuya representación se puede ver en la Figura (1).

Como se mencionó anteriormente, esta red corresponde a un grafo con peso, por tanto, las estadísticas típicas como el grado medio o clusterización se deben interpretar tomando en cuenta dicha particularidad. La red que se presenta tiene la particularidad de representar probabilidades, es decir, esta acotado entre 0 y 1. Por esta razón se van a mostrar las estadísticas comparando la red sin pesos, es decir, considerando la existencia de un enlace si el enlace de la red original es no nulo.

Estadística	Original	Sin peso
Número de nodos	5.880	5.880
Total enlaces	160.312,83	444.872
Grado medio	54,53	151,31
¿Está conectada?	No	No
Diámetro	–	6
Camino medio	–	2,34
Densidad	–	2,57%
Clustering (medio)	0,30	0,88
Clustering (global)	0,36	0,97
Asortatividad	0,67	-0,01

Tabla 1: Estadísticas calculadas de la red de licitaciones original y sin peso. No se calculó el diámetro por el poco sentido de asignarle un camino a una probabilidad, siendo un proxy inválido como métrica de distancia. La densidad no se calculó por el poco sentido de asignarle una densidad, siendo una densidad de densidades ya que los enlaces ya en sí poseerían una probabilidad intrínseca.

En la tabla (1) se puede observar la comparativa de la red. Al observar la red podemos hacer la comparativa del total de enlaces versus el peso total de los enlaces, siendo casi un tercio de los enlaces sin peso, es decir, a pesar de tener 444.872 enlaces, la intensidad (la probabilidad asociada) de dichos enlaces son mucho menores a 1. El grado medio es de 54, interpretándose como "Cada nodo tiene de media un peso total de 54,53". El diámetro de la red sin peso es de 6, es decir, siendo una red con un camino máximo muy pequeño, en comparación a la cantidad de nodos que tiene. El clustering medio y global varía sustancialmente entre las dos redes, esto es por la intensidad propia de los enlaces. El clustering es una métrica que mide la conectividad de los vecinos de los nodos, que sea de 0,30 nos dice que de media, existe un 30% de probabilidad que los vecinos de los nodos también estén conectados, aunque con la interpretación pesada, debe tomarse en cuenta la intensidad del enlace. La assortatividad nos muestra la tendencia de los nodos a conectarse con sus semejantes, es negativo si los nodos de bajo grado se tienden a conectar con aquellos de alto grado, y es positivo si se conectan nodos de poco grado con nodos de

^{*}j.illanesva@udd.cl

poco grado, y los de alto grado con los de alto grado. En este caso la asortatividad es positivo, por lo tanto, los nodos en este caso, se tenderían a agrupar en grupos semejantes, los hubs pequeños con los hubs pequeños y los hubs grandes con los hubs grandes.

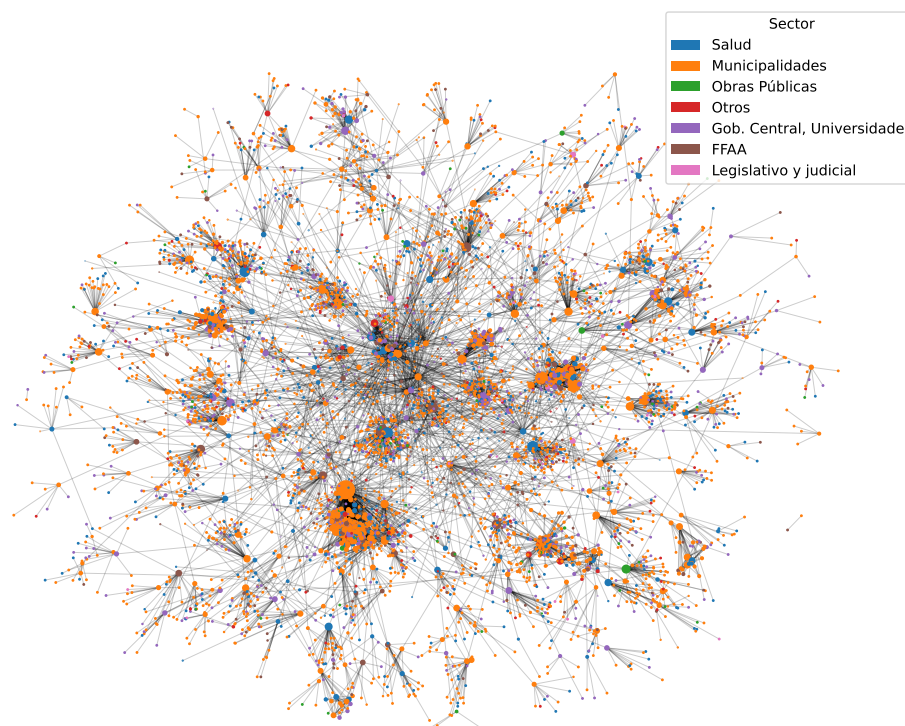


Fig. 1: Espacio de licitaciones. Se aprecia visualmente clusters de nodos y un comportamiento disasortativo. Principalmente las licitaciones corresponden a préstamo de servicios y adquisición de bienes para las municipalidades.

Las centralidades más importantes para investigar la red corresponden a las centralidades de grado y las centralidades de autovector. La centralidad de grado es importante para observar aquellos nodos (empresas) que poseen más influencia para ganar licitaciones, indicando algún tipo de colusión, corrupción o comportamiento subyacente que pueda indicar algún mecanismo externo. Sin embargo no se puede utilizar simplemente el grado, ya que zonas como el norte o el sur como se vio en el trabajo 1, tienden a no compartir empresas locales, generando un sesgo por aquellas empresas que operan por todo el país, por tanto, la centralidad de autovector permite mostrar la distribución estacionaria final de nodos que se llevarían la proporción más grande de licitaciones al iterar y ponderar sus conexiones. Se puede observar dichas visualizaciones en la figuras (3) y (4).

Si observamos el top 10 de ambas centralidades observamos una tendencia clara. De las 10 empresas, 9 corresponden a empresas de venta y reparación de equipos médicos, que son especializadas en venderles a centros médicos, los cuales constantemente requieren de adquisiciones y servicios para las maquinarias hospitalarias o los insumos médicos, la única empresa que no está en dicho rubro es una empresa inmobiliaria.

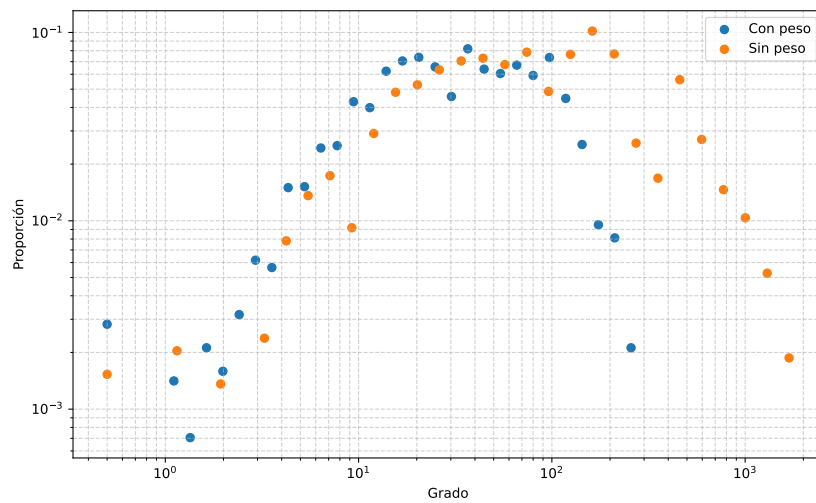


Fig. 2: Distribución de grado de las redes con peso y sin peso. La distribución de grado pareciera tener una curva en forma de U invertida, tanto para la red original como la red sin peso. Se debe aplicar un testeo de cola pesada para observar si existe una ley de escala tras un cierto umbral.

Top	Empresa
1	Imp. Dist. Arquimed Ltda.
2	B.BRAUN MEDICAL SpA
3	Inmobiliaria Innova SpA
4	ETHON PHARMACEUTICALS SPA.
5	Cencomex S.A.
6	Mediplex
7	FARMALATINA LTDA.
8	MADEGOM S.A.
9	Comercial LBF Limitada
10	Bluemedical SPA

Tabla 2: Top 10 empresas por centralidad de grado. La única empresa que no está asociada al sector médico (ya sea por insumos, reparación de implementos médicos, o fármacos) es Inmobiliaria Innova.

Top	Empresa
1	ATV ELECTRONIC
2	PIMET S.A.
3	ONESERVIS SPA
4	JUMACOBÁ
5	FRER ARQUITECTURA
6	ASI retail ltda.
7	TRI INGENIERIA
8	PLAGACCION
9	INSOLTEC
10	Roberto Lorca

Tabla 3: Top 10 empresas por centralidad de autovector. En este caso se observa más variación: Electrónica, Retail, Ingeniería, Control de Plagas, Ciberseguridad, entre otros.

2 Pregunta comparativa

Particularmente al analizar la red observando los estadísticos calculados anteriormente. Por ejemplo, tomemos en cuenta la red sin peso cuya densidad (que en caso de ser una red random corresponde con la probabilidad de enlace) es de 2,57%, teóricamente una red random tiene una clusterización igual a

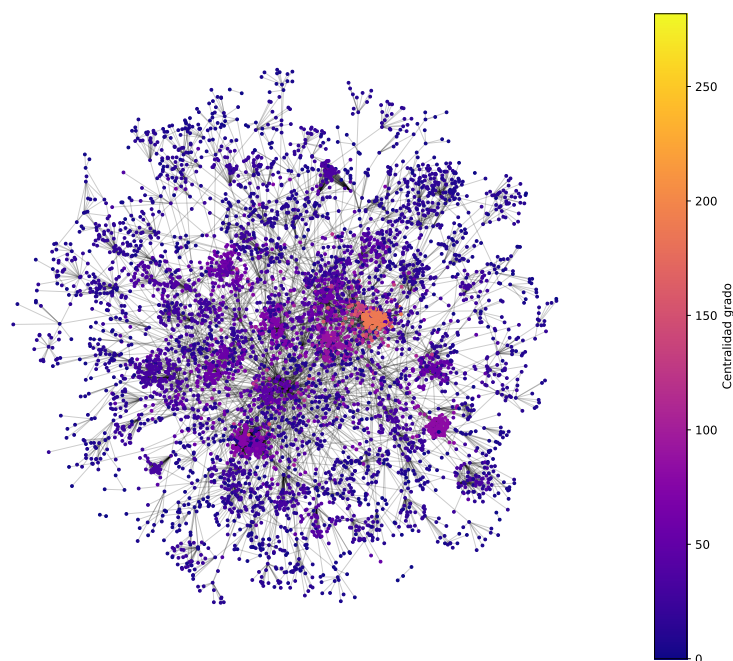


Fig. 3: Centralidad de grado. Es posible apreciar que la mayoría de los nodos poseen poca centralidad de grado, algunos una centralidad media, y un clúster con la mayoría de centralidad de grado, un guiño a la ley de escala. Se calculó utilizando la red original (con pesos).

su probabilidad de enlace, sin embargo, la clusterización obtenida es de un 90% aproximadamente, y si observamos la red pesada, de un 30%, difiriendo significativamente a la predicción teórica.

Por tanto, el análisis se restringe a una red Wattson–Strogatz o una Barabassi–Albert. Esto se confirma al observar el diámetro de la red, a pesar de tener miles de nodos, la distancia máxima entre dos nodos es de 6, razonablemente pequeño para ser una red sparse. Aunque, por el fenómeno que se está analizando, la probabilidad conjunta de las empresas de ganar licitaciones, debería (se hipotetiza) existir un preferential attachment por algunas empresas, por su tamaño, por la influencia política o económica, o por el tipo de servicio que brinda hacia el estado. Por tanto se testeara para las redes WS y BA.

3 Modelos nulos

Los parametros de control que tienen las redes de Wattson–Strogatz es la probabilidad de reconexión ya que se fija la cantidad de nodos y la cantidad de enlaces presentes, por lo tanto hay que buscar cual es la probabilidad de reconexión p optima para nuestra red. Para esto realizamos un gráfico p vs Clustering Global y la comparamos con nuestra red sin peso. Sin embargo como se puede observar en la figura (5), la clusterización obtenida de la red WS es muy baja comparada a la obtenida realmente. Es más, si observamos la distribución de grado obtenida por una red de WS, variando la probabilidad de reconexión (pero fijando el tamaño de la red y la cantidad de enlaces), la distribución de grado no se puede explicar por la red de WS, como se muestra en la figura (6). Por esta razón considero que la red de licitaciones **no** es una red de WS.

Para el caso de una red Barabassi–Albert también es complicado, dado que esta red posee una ley de escala con exponente $\gamma = -3$, por lo tanto, no tiene una variabilidad suficiente (más allá de fijar la cantidad de nodos y de enlaces) que permita explicar la distribución obtenida. Se realizó una test de ley de escala para la cola derecha de la distribución de grado de la red sin peso obteniéndose un exponente $\alpha = 1.11$ (Ajuste log-log, $p < 0.001$, $R^2 = 0.75$). Se puede apreciar la distribución de la red de licitaciones sin peso y la red de Barabassi–Albert en la figura (7). Por otro lado, la clusterización global calculada para la red sin peso es de 0.9, mientras que la obtenida por las redes BA es de 0,0686, por estas razones, considero que la red de licitaciones **no** es una red de BA, pero si posee una ley de escala no explicada por dicho modelo.

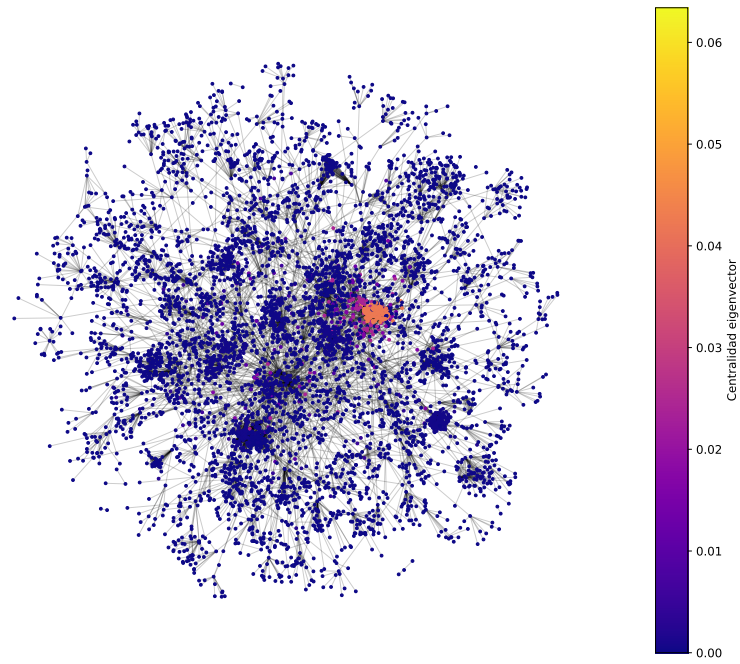


Fig. 4: Centralidad de autovector. Respecto a la centralidad de grado, se hace más evidente la centralización del clúster de color naranja, aunque también se puede apreciar la emergencia de ciertos nodos externo al clúster central.

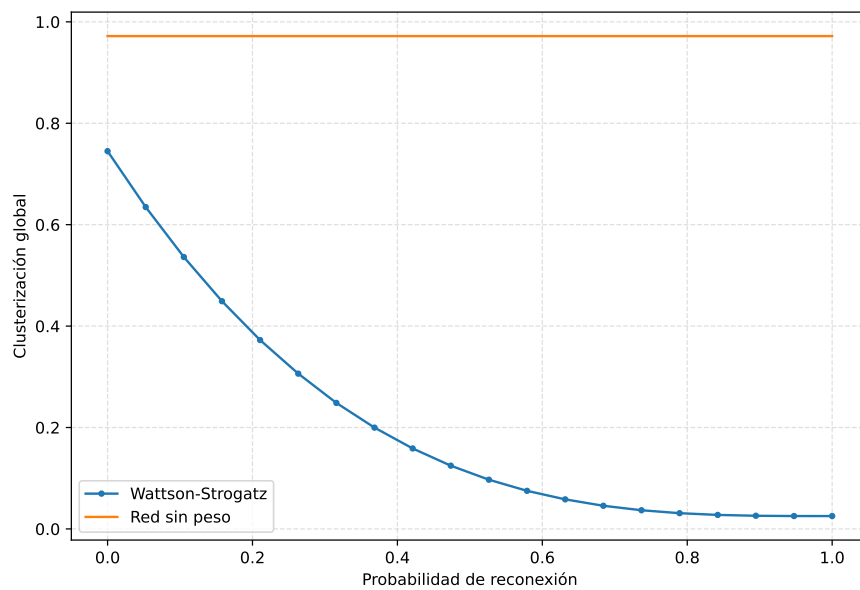


Fig. 5: Clusterización de la red WS versus la probabilidad de reconexión. Se aprecia que la clusterización es inversamente proporcional a la probabilidad de reconexión, pero comparado a la red sin peso, es demasiado bajo para ser explicado por este modelo.

4 Simulación

Para la generación de las figuras (5) y (6) se calculó una red WS con $N = 5.880$ y una conexión entre vecinos de $\langle k \rangle = 151$, se calcularon 20 redes por cada valor de probabilidad de reconexión p que fue variado secuencialmente en intervalos de 0.1, desde 0 hasta 1. Los histogramas fueron calculados

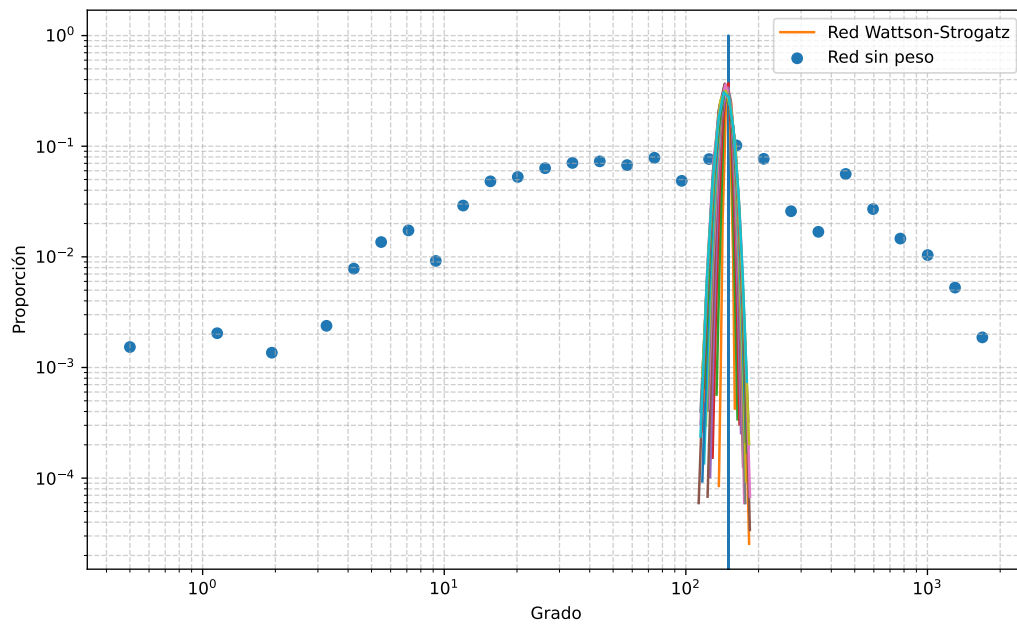


Fig. 6: Distribución de grado comparada con las distribuciones de grado de la red Wattson–Strogatz. La distribución de grado de la red modelada varía desde una probabilidad de reconexión de 0 a 1, pero la desviación estándar que surge es demasiado pequeña para explicar la distribución que se obtiene.

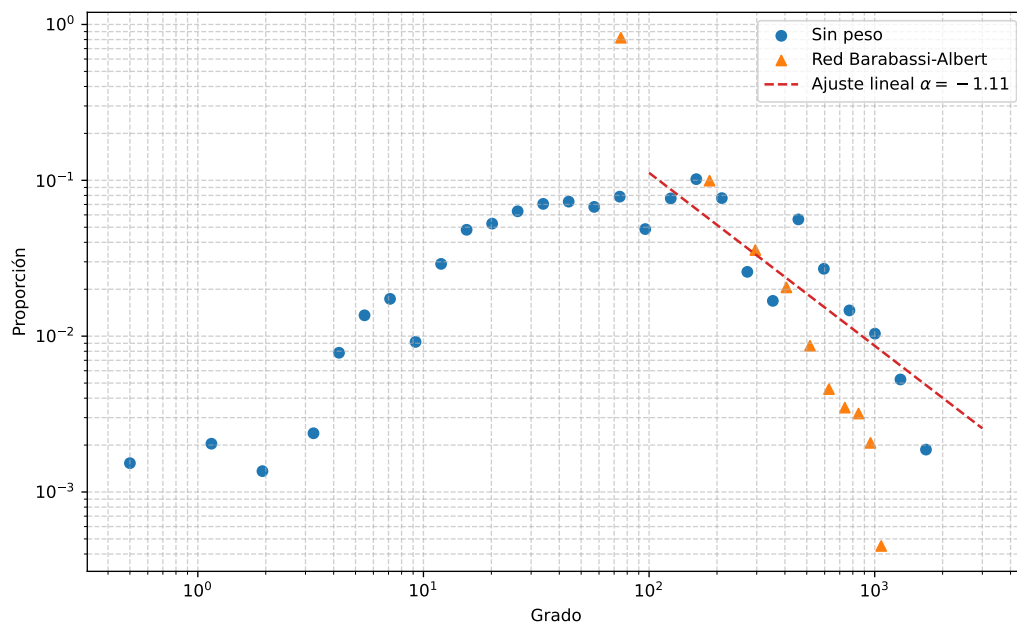


Fig. 7: Distribución de grado comparada con una red de Barabási–Albert. Se aprecia que la distribución de BA es una ley de escala pura, teóricamente es de un exponente de -3, y experimentalmente se puede apreciar que se sigue aproximadamente dicha predicción, sin embargo, la cola pesada de la distribución de la red sin peso no sigue exactamente dicho ajuste. Por otro lado, el exponente calculado por ajuste logarítmico es de -1.11, así como los grados medios y bajos tiene una curvatura hacia abajo.

utilizando la muestra de 20 redes.

Para el cálculo de la clusterización de la red BA y la figura (7) se calculó 20 redes de BA con un tamaño de $N = 5.880$ y una cantidad de reconexiones por paso de $m = \langle k \rangle / 2 = 75$

5 Conclusión

Por lo visto, la red no puede ser explicada por los modelos clásicos de la teoría de redes, esto es por dos motivos fundamentales: El primero es que la red original contiene peso, por tanto, es necesario describir un mecanismo generador que tenga en cuenta dicha particularidad. Lo segundo es que la distribución de grado obtenida no sigue las distribuciones de los tres modelos, es decir, ni una binomial, ni una exponencial ni una ley de escala con exponente -3. Naturalmente esto se debe a que no debe existir una razón restrictiva para que dicha ley ocurre cuando existe un preferent attachment ligado a los grados de los nodos, pero pueden existir otros parámetros que alteren dicha estructura, como se mencionaba anteriormente, corrupción, colusión, o especialización en distintos rubros. Por ejemplo, se vio que el mayor grado se lo llevan empresas asociadas al rubro de la salud, al ser un sector público que requiere de constantes licitaciones para abastecerse y reparar equipos urgentemente, sin embargo, al observar la centralidad de autovector, el algoritmo asocia como importantes empresas más variadas o técnicas, esto es algo que el modelo de BA no captura al desechar dichos atributos solo enfocándonos en la propia conexión de los nodos.

Particularmente, este trabajo se centró en la red de licitaciones especializadas y ganadoras de concursos públicos en enero del 2025, pero mi interés se centra en dos aspectos extra: La *temporalidad* de dichas licitaciones, es decir, observar si esta estructura vista es estática o dinámica, y las razones del porqué, además de escudriñar al respecto del porqué tiene dicha distribución, quizás, buscando un modelo que tome en cuenta esta distribución en forma de U invertida en log-log (que por lo demás, hipotetizo que es una distribución hiperbólica por tener dos colas pesadas). El otro aspecto es considerar el contrafactual de la red, en este caso, vimos la red de licitaciones **ganadas** pero perdemos la información de las licitaciones perdidas, la pregunta nuevamente es ¿Por qué? ¿Es la estructura de la red misma de licitaciones ganadas? ¿Se debe a la información implícita que se mueve por la red? ¿Es acaso el precio ofertado de las licitaciones?