

Intelligent Algorithms for Reducing Query Propagation in Thematic P2P Search

Ana L. Nicolini
Carlos M. Lorenzetti
Ana G. Maguitman
Carlos I. Chesñear



Grupo de Investigación en
Recuperación de Información y
Gestión del Conocimiento.

Universidad Nacional del Sur
Av. L.N. Alem 1253
Bahía Blanca - Argentina



Contenido

- Problemática
- Marco teórico
- Algoritmos propuestos
- Implementación
- Simulaciones y estadísticas
- Trabajo a futuro
- Conclusiones



Problemática



Problemática

- Los buscadores que trabajan en entornos distribuidos se enfrentan a los problemas de cobertura y sensibilidad al contexto, ocasionando un deterioro en la en la precisión de las respuestas.
- La situación se ve agravada dado que en este tipo de entornos la información se encuentra esparcida a través de diversos nodos totalmente independientes entre si .



Problemática

- A raíz de esto surge la necesidad de desarrollar un sistema capaz de facilitar la búsqueda contextualizada y distribuida que a su vez controle la cantidad de consultas que circulen por la red.
- Otro requisito de este sistema es que en primera instancia no puede conocer dónde la información se encuentra almacenada ni con qué vocabulario se encuentra indexada.



Marco Teórico



Marco teórico

- Una *comunidad semántica* es un conjunto de nodos que comparten determinados intereses.
- Para evaluar el surgimiento de comunidades semánticas en nuestro trabajo utilizamos el concepto de *topología de mundo pequeño* y las métricas que determinan la existencia del mismo.

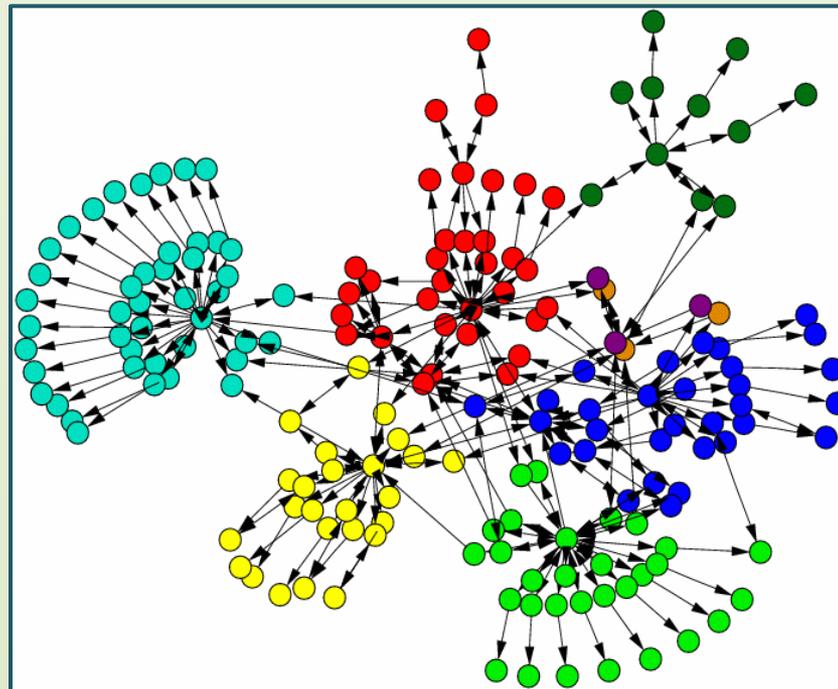


Marco teórico

- El coeficiente de clustering de un nodo indica qué tan interconectado está con sus vecinos.
- Un valor máximo indica la formación de un clique (grafo completo).
- Un valor pequeño indica que el nodo está bajamente conectado en la red.
- El promedio de este coeficiente entre los nodos de la red, en el marco de nuestro trabajo, nos permite cuantificar el surgimiento de las comunidades semánticas.

Marco teórico

- Se dice que una red posee una *topología de mundo pequeño* cuando el promedio de la longitud de pasos entre dos nodos de la misma es pequeño y el coeficiente de clustering es alto.





Algoritmos Propuestos



Algoritmos propuestos

- Propusimos dos algoritmos de aprendizaje gradual que comparten las siguientes características:
 - El algoritmo se ejecuta de forma totalmente independiente en cada nodo.
 - Se almacena en una tabla el conocimiento adquirido, ej:

Química#_22#_46#_99
Matemática#_2#_62#_899



Algoritmo adaptativo

- Inicialmente el nodo no conoce nada respecto a su entorno. Por esta razón, al momento de realizar una consulta se envía la misma a todos sus nodos adyacentes.
- Cuando llega un mensaje de consulta que se puede contestar, se genera el mensaje de respuesta. La particularidad es que cada nodo que se encuentra en el camino de la respuesta aprenderá en qué temática está especializado el nodo que contestó.

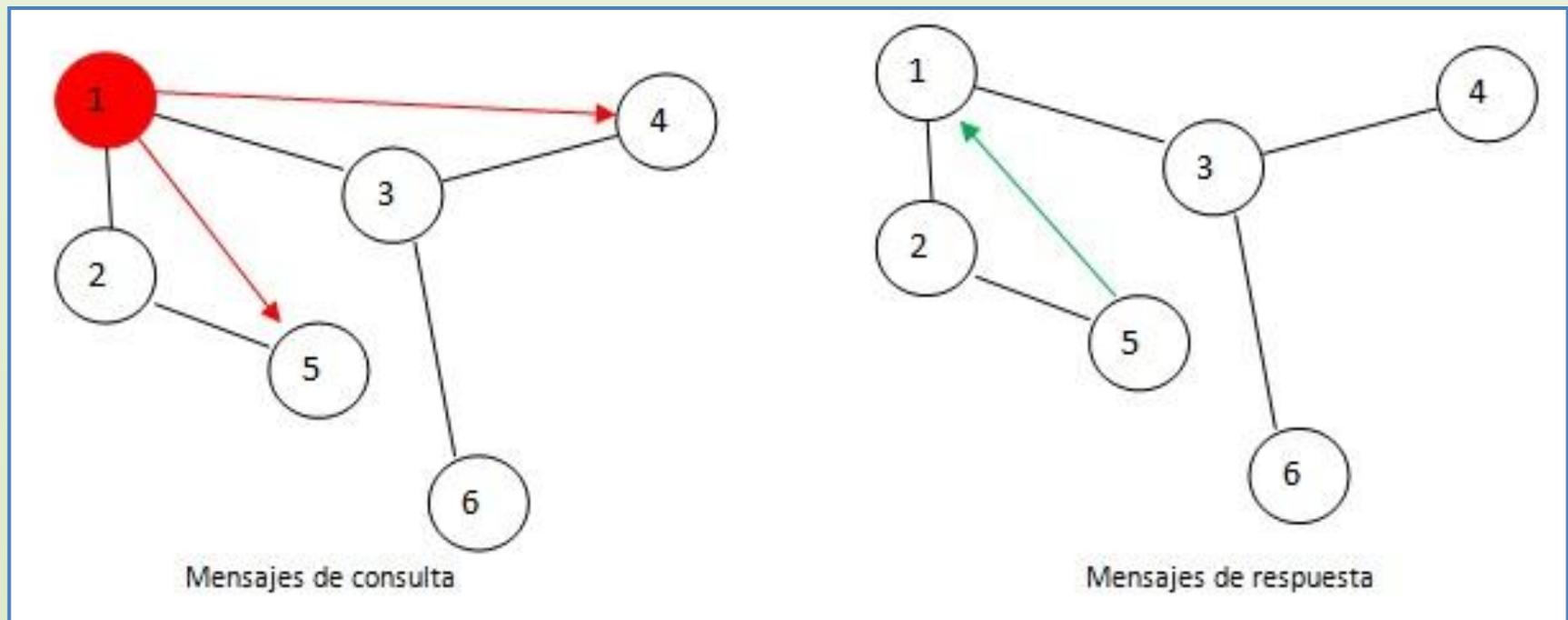


Algoritmo adaptativo

- La consecuencia directa de este aprendizaje es que al momento de realizar una nueva consulta sólo se la enviará a aquellos nodos que estén relacionados con la temática de la misma.
- Veremos con un ejemplo el comportamiento de este algoritmo en distintas situaciones.

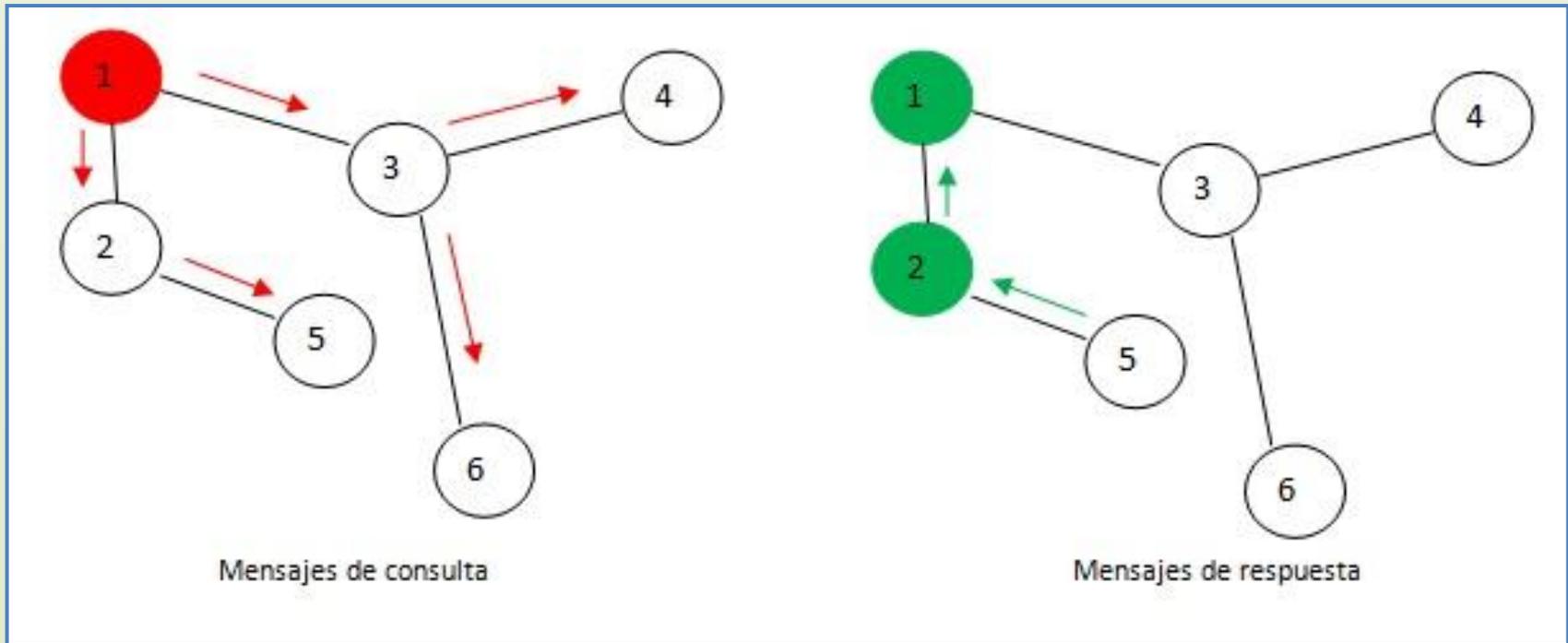
Algoritmo adaptativo

- El nodo 1 quiere hacer una consulta que puede resolver nodo 5. Sabe que los nodos 4 y 5 tienen su misma temática:



Algoritmo adaptativo

- En este caso el nodo 1 no tiene en su tabla TN a ningún nodo relacionado con la temática de la consulta:





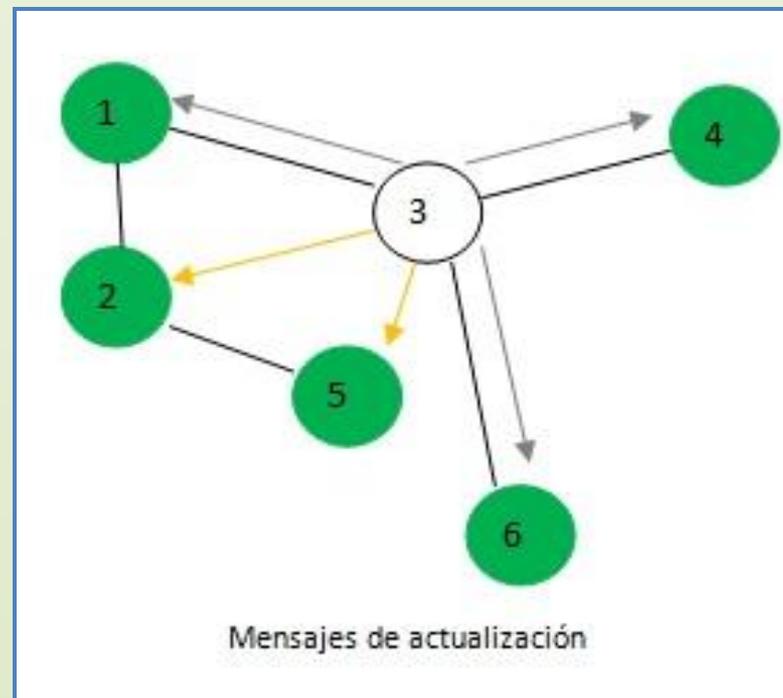
Algoritmo adaptativo

- Además de los mensajes de consulta y de respuesta existen los mensajes de actualización.
- Estos mensajes se utilizan para incrementar el conocimiento global de la red.
- Este tipo de mensajes lo envía un nodo a sus adyacentes y a los de su comunidad cuando agrega nuevo conocimiento a su tabla.



Algoritmo adaptativo

- Supongamos que el nodo 3 tiene la temática X y recibe un mensaje de respuesta sobre la temática X. Supongamos también que el nodo 3 sabe que los nodos 2 y 5 pertenecen a su comunidad:



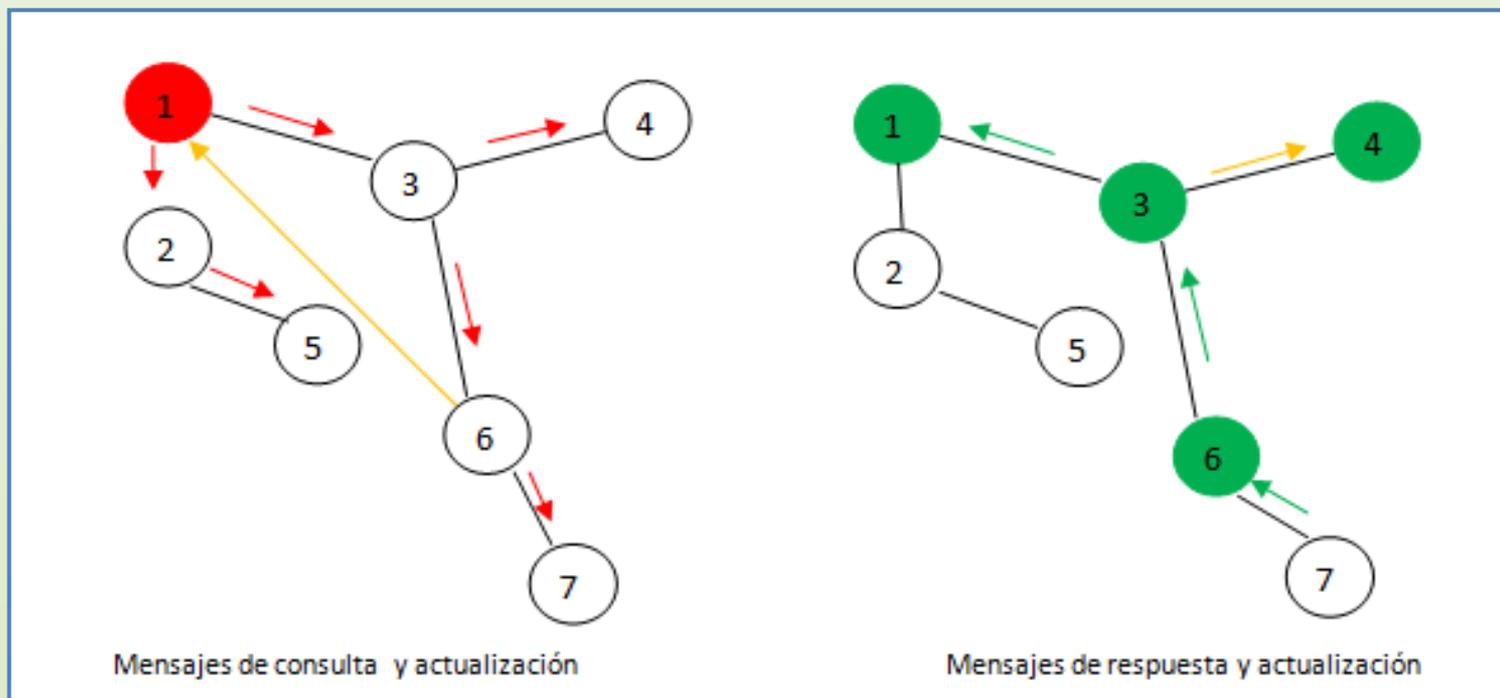


Algoritmo adaptativo

- Existe un último caso en el que se envía un mensaje de actualización.
- Cuando a un nodo le llega un mensaje de consulta por broadcast y la temática de la misma coincide con alguna de las temáticas del nodo, y además, no puede responder, entonces envía un mensaje de actualización al nodo que originó la consulta.

Algoritmo adaptativo

- Supongamos que el nodo 1 genera una consulta a la que puede responder el nodo 7, pero 1 no tiene conocimiento de 7. El mensaje de consulta llega al nodo 6 (que tiene al menos una temática coincidente con 1):



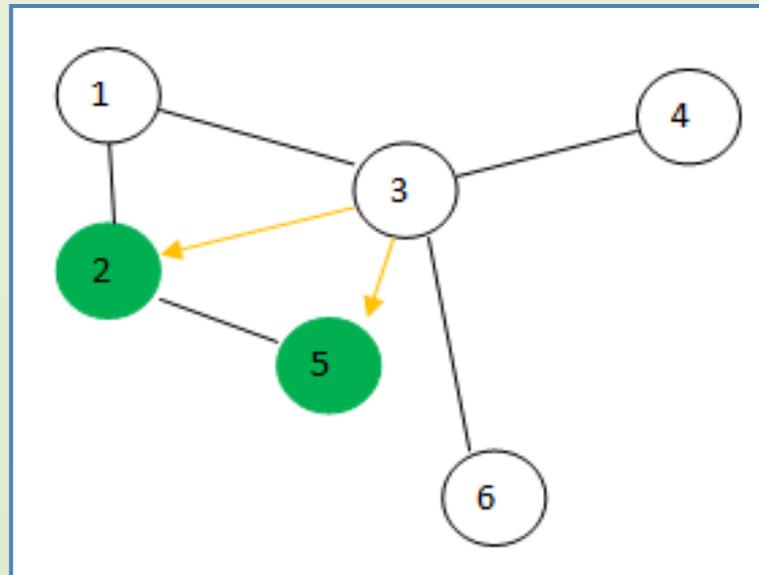


Algoritmo adaptativo selectivo

- Similar al algoritmo anterior excepto que se omiten los mensajes de actualización a los nodos adyacentes.
- Sólo se envía este tipo de mensajes en los casos que se muestran a continuación.

Algoritmo adaptativo selectivo

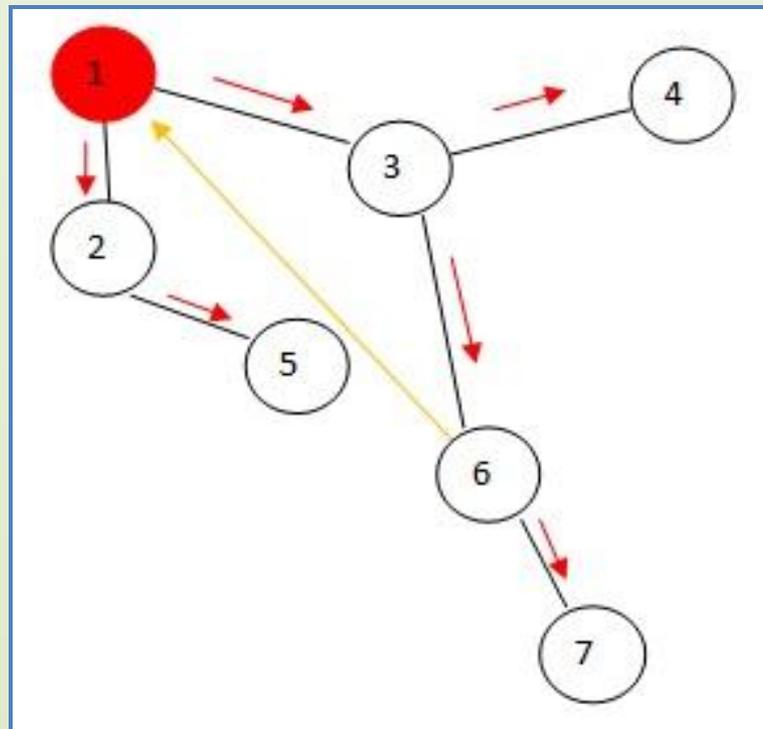
- Que a un nodo llegue un mensaje de respuesta cuya temática coincida con algunas de las del nodo, entonces se envía un mensaje de actualizaciones a todos los miembros de la comunidad a la que pertenece el nodo en cuestión.





Algoritmo adaptativo selectivo

- Que a un nodo le llegue un mensaje de consulta por broadcast y la temática de la misma coincide con alguna de las temáticas del nodo, entonces envía un mensaje de actualización al nodo que originó la consulta.



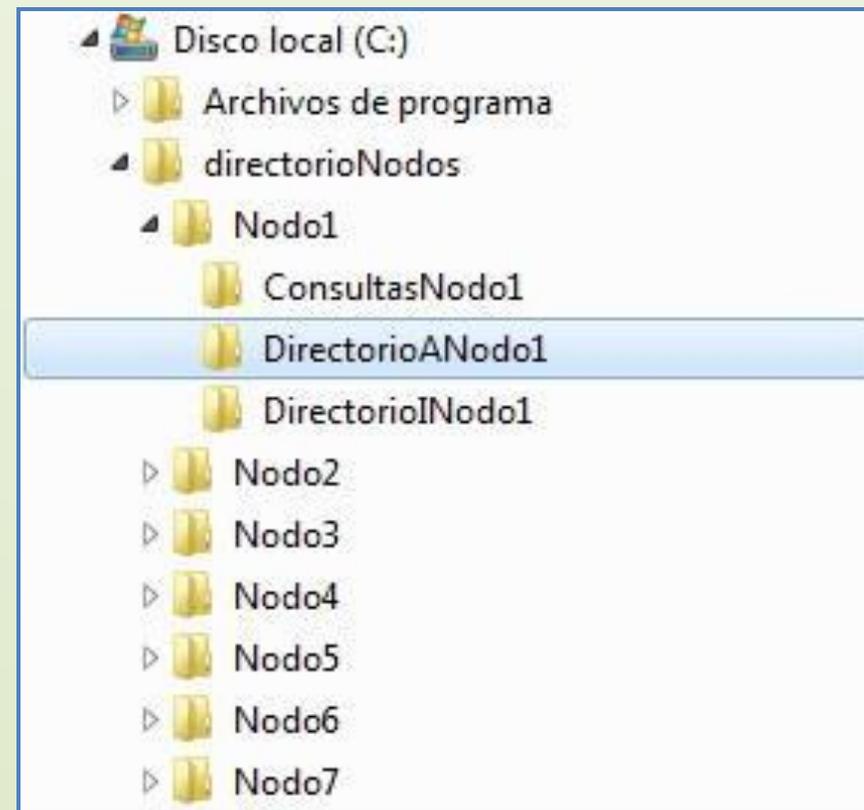


Implementación



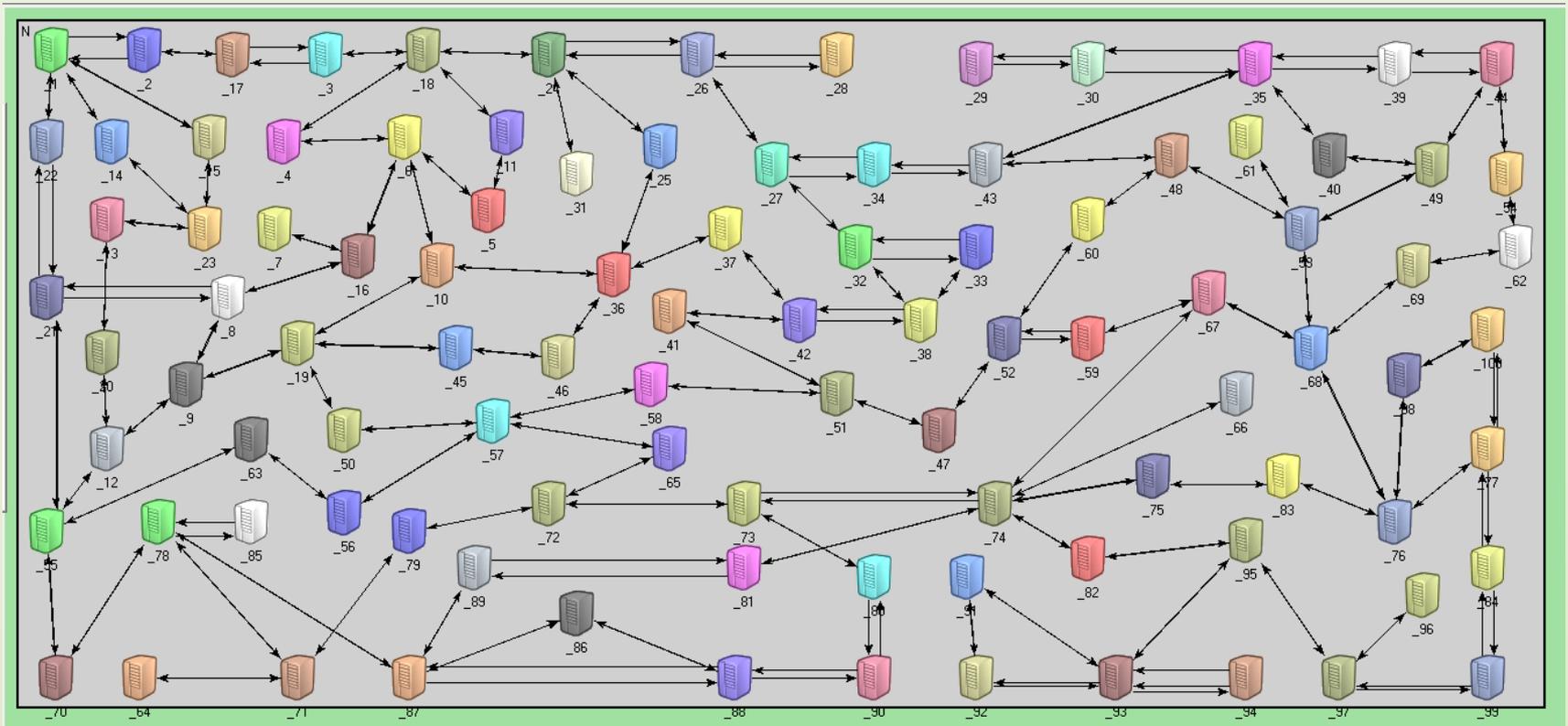
Implementación

- 40000 artículos científicos
- 13 temáticas
- 1000 nodos.



Implementación

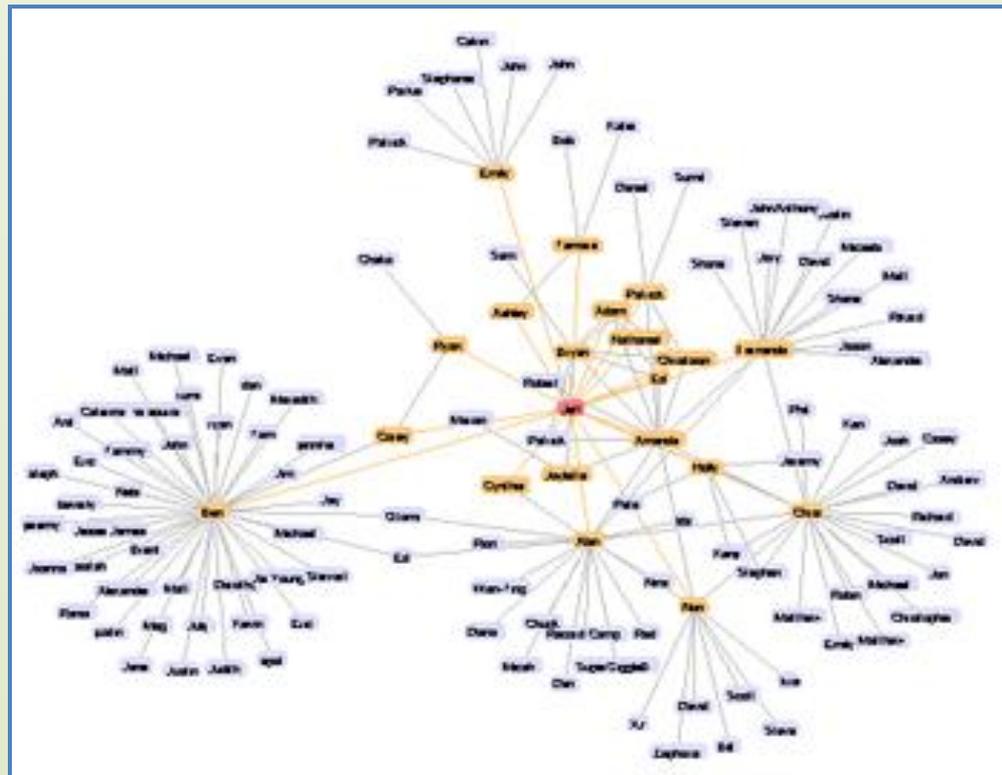
- Framework OmNet++ (Versión Java)



OMNeT++

Implementación

- Para la visualización de la red lógica se utilizó la herramienta JUNG (Java Universal Graph)





Simulaciones y Estadísticas

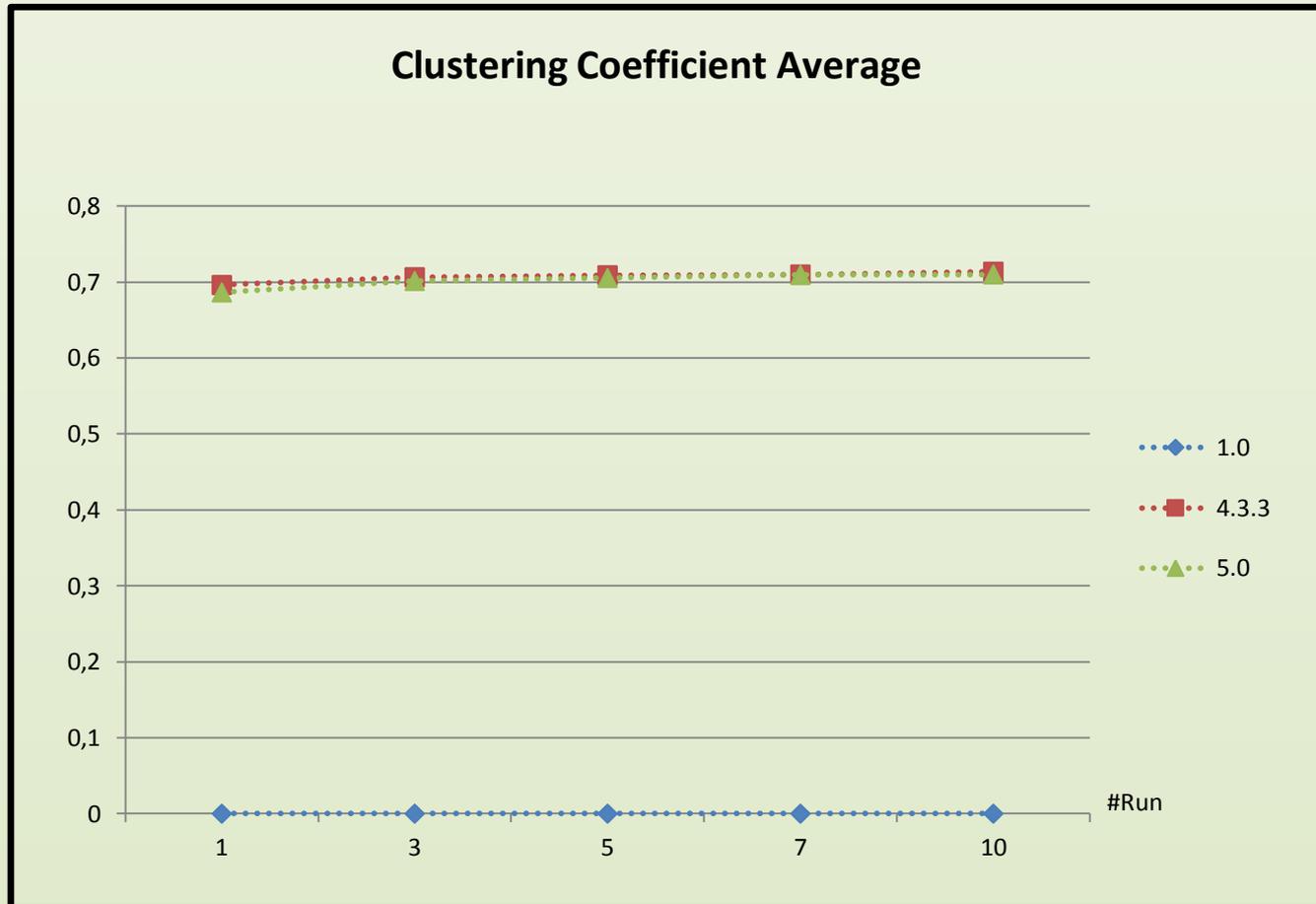


Simulaciones y estadísticas

- Como en nuestro abordaje trabajamos con algoritmos con aprendizaje incremental, se realizaron 10 corridas con cada uno para poder obtener datos significativos.
- A modo comparativo también incluimos en la simulación un algoritmo sin inteligencia alguna, al estilo Gnutella.

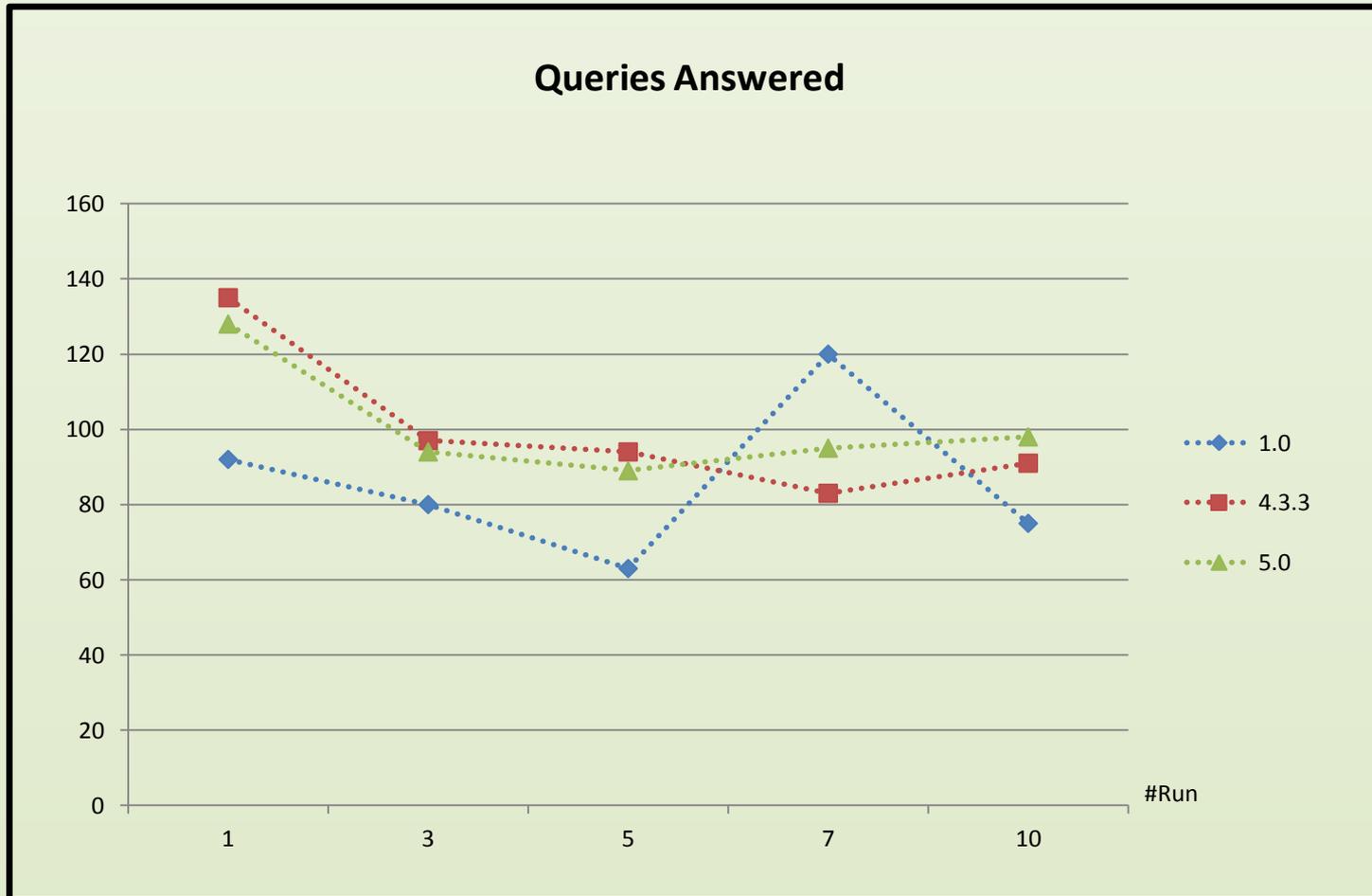


Simulaciones y Estadísticas



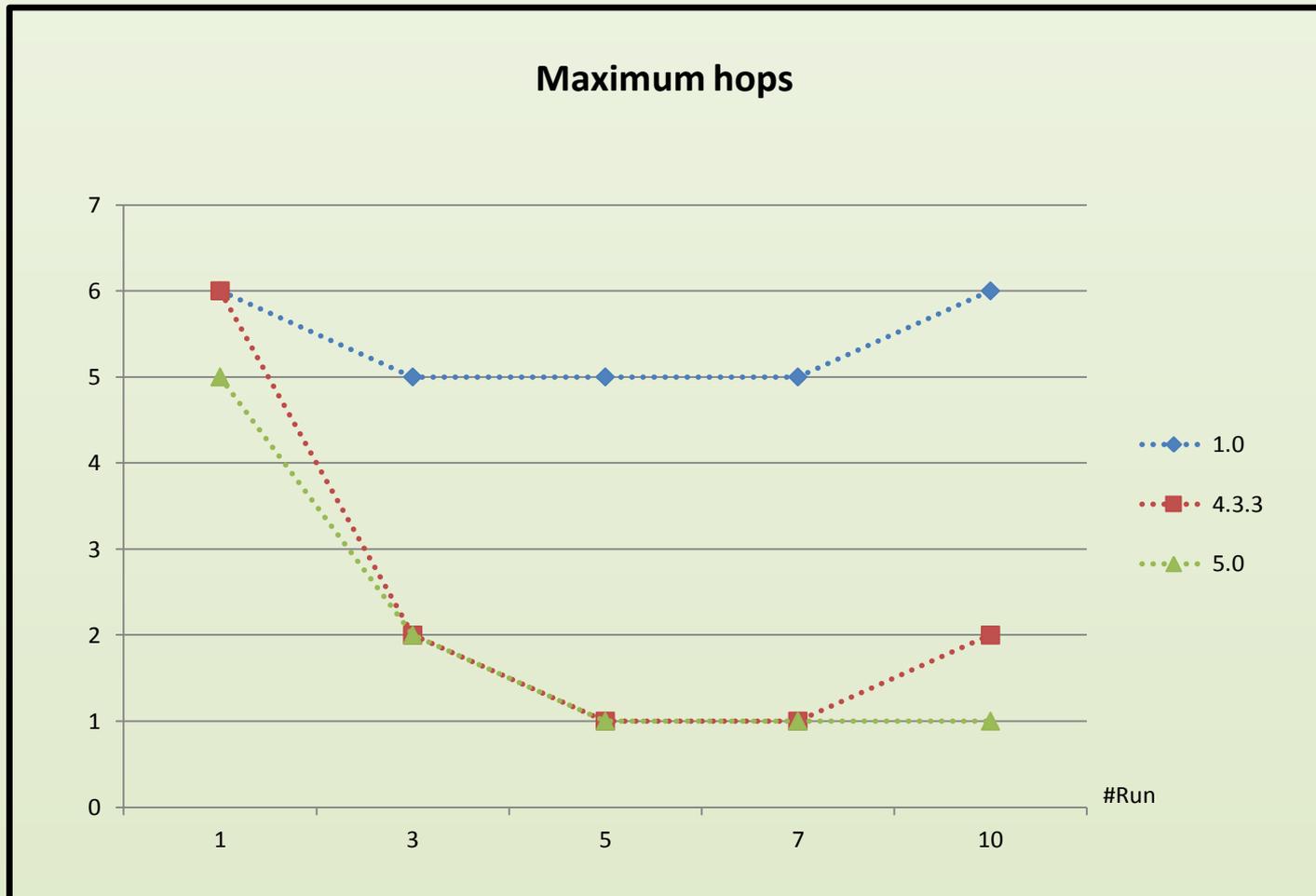


Simulaciones y Estadísticas





Simulaciones y Estadísticas

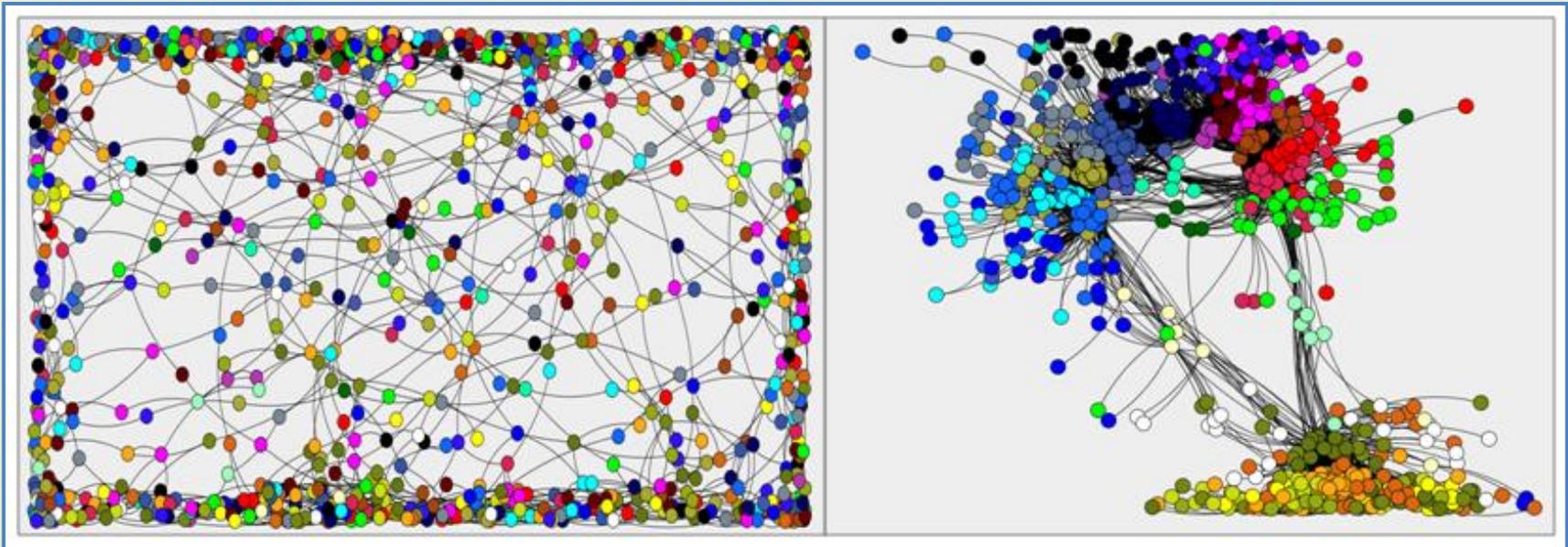




Simulaciones y Estadísticas



Simulaciones y Estadísticas





Trabajo a Futuro



Trabajo a futuro

- Mejorar la sensibilidad al contexto. Lo ideal sería inspeccionar la consulta y mediante un clasificador determinar a qué temática pertenece, para luego poder enviar el mensaje a los nodos correspondientes.
- Implementar un sistema con búsqueda incremental de buenos descriptores y discriminadores en el contexto temático de los usuarios, donde la consulta puede refinarse valiéndose del vocabulario aprendido.



Trabajo a futuro

- Realizar la búsqueda de los archivos de respuesta no sólo por sintaxis sino también por semántica.
- Implementar mecanismos de curiosidad para solucionar el problema de las comunidades cerradas.
- Para obtener un escenario completo, los nodos deberían generar consultas dinámicamente y poder ocasionalmente cambiar de intereses.



Conclusión



Conclusión general

- Corroboramos nuestra suposición inicial de que cuando los nodos transmiten el conocimiento que adquieren hacia el resto de la comunidad aumenta el grado de localidad semántica.
- El aprendizaje no sólo se efectúa al determinar qué nodo responde una consulta, sino también tras encontrar similitud semántica entre el nodo demandante y el nodo receptor (o intermediario), independientemente si puede satisfacer o no la consulta.



Conclusión general

- Asimismo estos algoritmos demostraron que se logra formar una red lógica que presenta una topología de mundo pequeño, evidenciando así el surgimiento de comunidades semánticas.
- El tiempo de procesamiento que requieren los algoritmos inteligentes no produce un incremento significativo en el tiempo de respuesta.
- Se debe moderar el envío de mensajes de actualización. Estos mensajes permiten aumentar la velocidad de aprendizaje de la red pero al mismo tiempo la congestiona, ocasionando pérdidas de mensajes y aumentando el tiempo de respuesta.

Intelligent Algorithms for Reducing Query Propagation in Thematic P2P Search

Ana L. Nicolini
Carlos M. Lorenzetti
Ana G. Maguitman
Carlos I. Chesñear



Grupo de Investigación en
Recuperación de Información y
Gestión del Conocimiento.

Universidad Nacional del Sur
Av. L.N. Alem 1253
Bahía Blanca - Argentina