

Mineria de Grafos en Redes Sociales usando MapReduce

Jose Gamez¹ and Jorge Piloza¹

Carrera de Ingeniería en Sistemas Computacionales
Universidad de Guayaquil

1. Introduccion a la Problematica

Recordemos que en la actualidad estamos viviendo en la era o en la epoca de los datos. Vivir en esta epoca ha sido gracias a los avances de las tecnologias de la informacion y de la comunicacion que nos permite coleccionar y almacenar informacion de actividades comerciales; academicas; cientificas y de la vida cotidiana de cada individuo de esta epoca; que se lo realiza de manera facil y economica la recoleccion de la informacion anteriormente mencionada.

Tambien debemos destacar o recalcar que el creciente desarrollo paulatino de la tecnologia en el campo de la educacion en la actualidad; esta contribuyendo en la mejora del aprendizaje academico del estudiante; donde se hace mas notable en la Educacion a Distancia; cuando se crearon las herramientas que facilitan la comunicacion y la creacion del trabajo grupal; adicional con el uso de metodologias y tecnicas que permite evaluar el uso de dichas herramientas mencionadas; proporcionando un relevante y significativo conocimiento con el descubrimiento de patrones que sirve para la toma de decisiones y por lo consiguiente, con el planteamiento de estrategias que nos permitan mejorar el proceso de enseñanza-aprendizaje entre los alumnos y el profesor que cada materia que se este estudiando.

A nuestro grupo nos toca el tema a exponer sobre la Minería de Grafos en redes sociales; donde no hemos abarcado durante todo el curso o le hemos orientado en el area de la educacion universitaria para ser mas especificos nuestro grupo lo hemos enfocado para la Universidad de Guayaquil de la carrera de Ingenieria en Sistemas Computacionales y Networking con sus respectivos alumnos o estudiantes de cada una de las carreras anteriormente mencionadas.

2. Antecedentes

El campo que vamos abarcar sobre los antecedentes de Minería de Grafos de Redes Sociales; es como en años anteriores los estudiantes de las carreras de Ingenieria en Sistemas Computacionales y Networking como han mejorado sustancialmente sus notas academicas o evaluaciones academicas desde se introdujo la famosa y potente red social llamada Facebook; se debe recalcar en los inicios de nuestra nuestra era dificil el acceso al internet por obvias razones conocidas.

3. Motivacion

Es sobre el estudio de tecnicas para solucionar lo siguiente:

- Obtener los nodos mas importantes

TEMA		Todos los días	1/2 a la semana	Acto Importante	Casi Nunca	Nunca	Total
Amigos		1	8	23	8	0	40
	%	33,3%	53,3%	63,9%	34,8%	0%	51,30%
Familia		1	0	0	0	0	1
	%	33,3%	0 %	0 %	0 %	0%	1,30%
Estudios		0	0	2	5	1	8
	%	0 %	0 %	5,60%	21,7%	100%	10,30%
Afición		1	7	11	10	0	29
	%	33,3%	46,7%	30,60%	43,5%	0%	37,2%
Total		3	15	36	23	1	78
	%	100%	100%	100%	100%	100%	100%

- Determinar la influencia de estos nodos sobre el grafo
- Encontrar los patrones cercanos para alcanzar el maximo rendimiento

4. Datos, Gráficas o Análisis Estadísticos de las Variables

Se ha planteado durante todo el curso las siguientes variables que hemos estado usando en el tema de la Minería de Grafos de Redes Sociales que son:

- Planteamiento y resolución de dudas sobre la asignatura
- Materias más buscadas por su difícil comprensión o temas no comprendidos en clase

En el primer punto que se trata sobre el Planteamiento y resolución de dudas sobre la asignatura; quiere decir sobre las dudas puntuales que un estudiante de cualquiera de las dos carreras tienen sobre una materia en específico como por ejemplo: cuando uno estudiante no comprendió en clases el tema de Herencia de Clases en Java y entre otros ejemplo similares que podrían presentarles al estudiantado.

El segundo tema que vamos a tratar es Materias más buscadas por su difícil comprensión o temas no comprendidos en clase; nos referimos a que materias que ha visto o esta viendo algún estudiante de cualquiera de las dos carreras se le ha complicado comprenderla; ya que puede presentar varias circunstancias que se le complicaría al estudiante; sea por el grado de complejidad de la misma materia; por falta de tiempo; por inasistencia a la curso; por cruce de materias entre otras alternativas que se le puede presentar al estudiante.

Temática fotos y frecuencia

Frecuencia de consulta del grupo de Facebook de la asignatura

Percepción de mejora del estudiante gracias al grupo de Facebook

5. Metodología

Las metodologías, técnicas o algoritmos que estamos usando son los siguientes: El primer algoritmo que hemos usado para realizar el emparejamiento de patrones o elementos característicos (repetitivos) es el método o algoritmo Subdue y con los resultados que nos dio el Subdue al encontrar los patrones repetitivos; aplicaríamos el algoritmo MapReduce

A continuación daremos una pequeña introducción o explicación de que se trata la metodología Subdue; se conoce como un Sistema de Aprendizaje entre relaciones que es usado para

Consulta grupo de Facebook	Porcentaje válido
Todos los días	42,3%
Una o dos veces a la semana	44,9%
Sólo cuando tengo que entregar un trabajo	12,8%
Total	100%

Opinión estudiante y mejora asignatura	Porcentaje válido
Mucho	75%
Poco	21,1%
Nada	3,9%
Total	100%

encontrar subestructuras que aparecen o se presentan repetidamente en la representación gráfica basada en los grafos dentro de una base de datos.

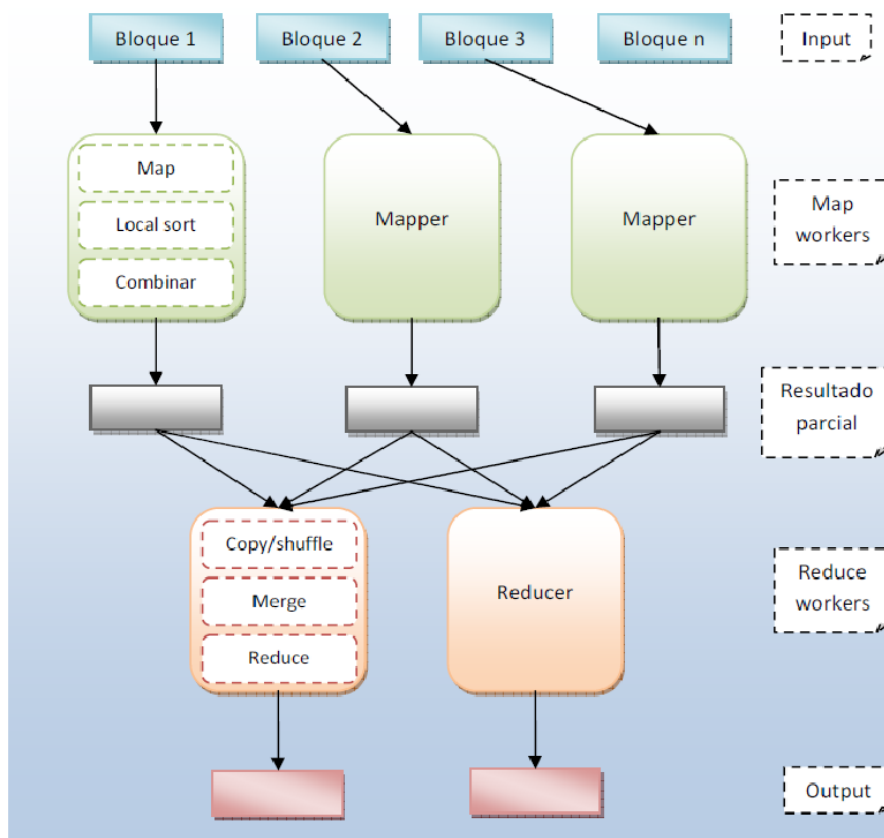
Subdue tiene la capacidad de realizar un proceso interno llamado macheo inexacto que es el encargado de encontrar o descubrir las subestructuras con pequeñas variaciones. Adicional el Subdue también usa un algoritmo de búsqueda. Donde el algoritmo de búsqueda consta con dos principales componentes que son :

- Espacio de búsqueda
- La estrategia de control

El algoritmo MapReduce es conocido en el mundo de la informática como un ambiente y un ambiente de programación desarrollado por la empresa de Google; donde su principal funcionalidad es de procesar grandes volúmenes de datos; donde este procesamiento se lo realiza de manera en párelo en un clúster compuesto por computadoras u ordenadores de uso general. Debemos destacar que el MapReduce no requiere una programación en paralelo por parte del usuario. MapReduce brinda las siguientes ventajas:

- Mayor escalabilidad (soporta miles de nodos)
- Tolerancia a fallas a gran escala.
- Flexibilidad en el manejo de datos no estructurados

Las funciones Map y Reduce se ejecutan de manera en paralela y distribuida en un cluster. Replica las funciones Map y Reduce en los nodos del cluster; de tal manera que la replicación de las dos funciones se ejecuten al mismo tiempo en nodos distintos



Interdependencia positiva	Construcción de Significado	Relaciones
Responsabilidad Individual	Explica / argumenta	Refuerza / aprueba
Propuesta de método	Reformulación / síntesis	Estimula / anima
Interpela Responsabilidad	Pregunta contenido / opinión	Agradecimiento
Preguntas del método	justifica	Dialogo social
Aclara / complementa organiza	Discrepa	Comunicación abierta
	Aclara / complementa contenido	
	Enunciados metacognitivos	
	Explica la tarea	

6. Resultados

Daremos una pequeña desventajas que hemos visto al usar la metodología MapReduce que podría afectar nuestros resultados que estemos buscando en la Minería de Grafos en Redes Sociales.

1. NO hay lenguaje de alto nivel.El usuario debe decodificar sus operaciones en las funciones Map y Reduce.
1. No tiene esquemas ni índices
1. Flujo de datos fijos
1. Baja eficiencia

En el siguiente esquema se muestra una tabla de indicadores de aprendizaje cooperativo

El desarrollo de la investigación inicia con una inspección estadística del uso general de las herramientas colaborativas en el entorno virtual, con el fin de tener una proyección de los datos a cuantificar por las diferentes técnicas de análisis. A continuación se muestra una contextualización del problema, describiendo el entorno y las herramientas

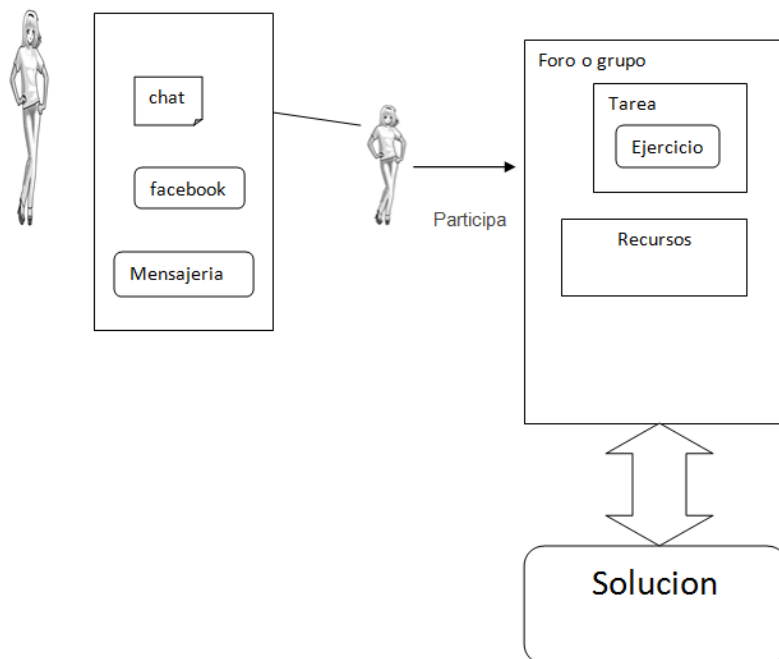
Clasificación de los medios de colaboración

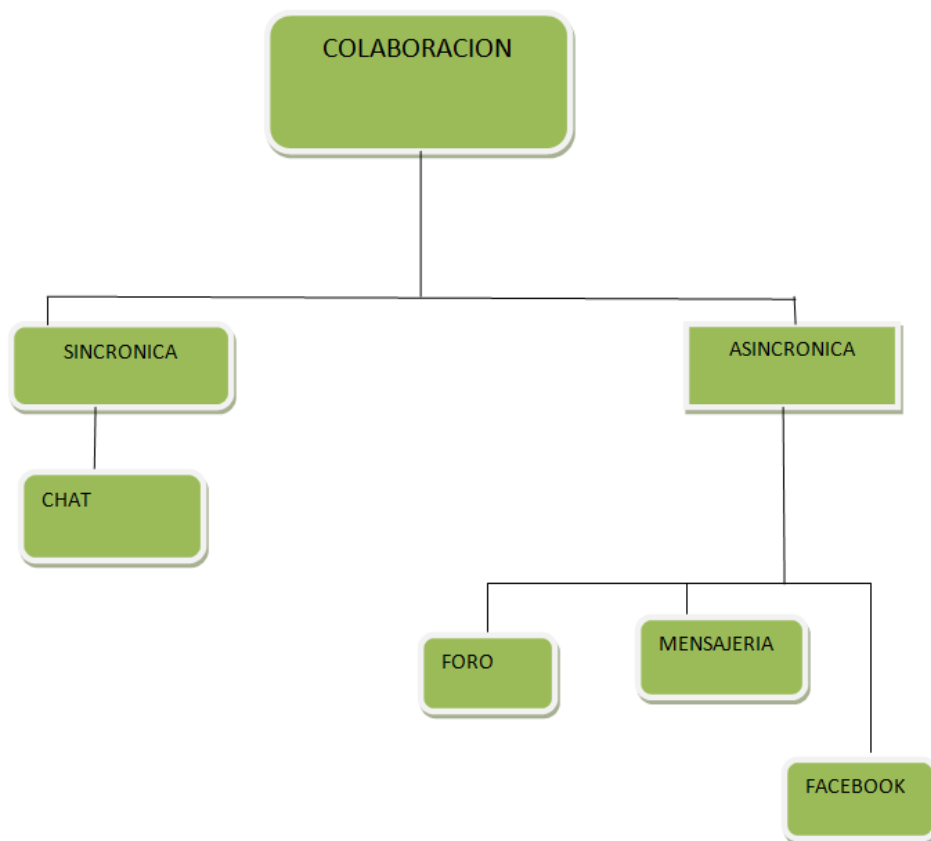
En la siguiente tabla y su respectiva grafica donde se muestra el numero total de interacciones en el foro

En la siguiente tabla y su respectiva grafica donde se muestra de usuario participativos en el foro

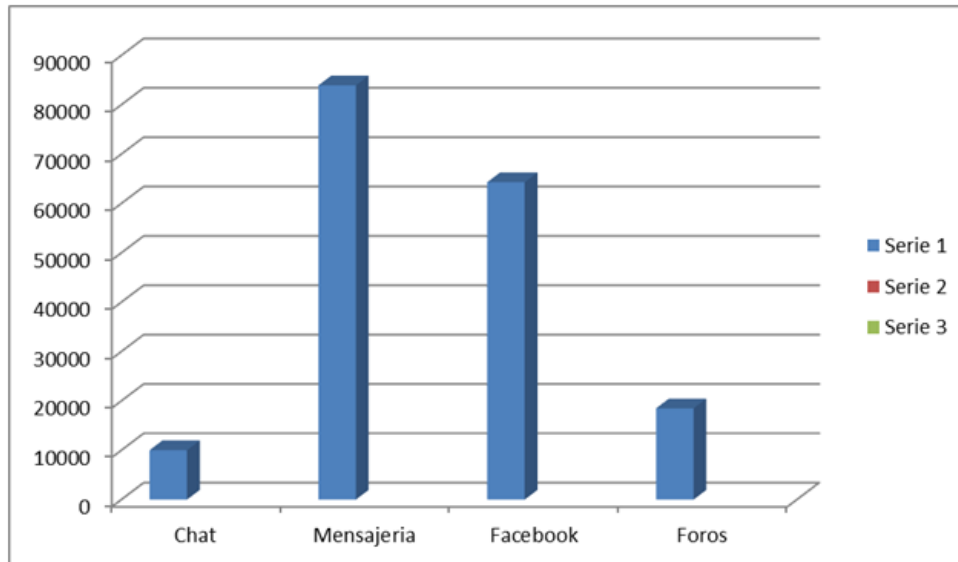
7. Bibliografia

<https://jcasasr.files.wordpress.com/2010/12/tfm-jordi-casas.pdf>
<http://fernandosantamaria.com/blog/tag/analisis-de-redes-sociales/>
<http://docplayer.es/3987429-Contenido-septiembre-2014.html>
<https://unpocodejava.wordpress.com/category/bigdata/page/9/>





<i>Herramientas</i>	Numero total de interacciones
Chat	10010
Mensajería	83925
Facebook	64301
foros	18495



Herramientas	Usuarios
Chat	255
Mensajería	13300
Facebook	11880
Foros	5237
Total usuarios	30109

