

TABLA DE CONTENIDOS

	página
Dedicatoria	I
Agradecimientos	II
Tabla de Contenidos	III
Índice de Figuras	VI
Índice de Tablas	VIII
Resumen	XI
1. Introducción	12
1.1. Descripción del problema	12
1.2. Hipótesis	15
1.3. Objetivos	15
1.3.1. Generales	15
1.3.2. Específicos	16
1.4. Metodología	16
1.5. Alcances	17
2. Contexto	18
2.1. Grafos	18
2.2. Sistemas de gestión de bases de datos de grafos	21
2.3. Benchmarks para bases de datos de grafos	26
2.4. <i>GDBenchmark</i>	29
2.4.1. Generador de datos	29
2.4.2. Generador de operaciones	30
2.4.3. Evaluador de operaciones	31

3. Métodos de selección de datos de prueba	33
3.1. Datos de prueba comparables	33
3.2. Caso de estudio	34
3.3. Análisis de consultas de prueba	35
3.4. Experimento modelo	41
3.4.1. Descripción de los experimentos	42
3.4.2. Criterios de selección de datos de entrada	43
3.4.3. Resultados empíricos	47
3.4.4. Análisis de los resultados	59
3.5. Método de selección de datos de prueba	64
4. Desarrollo del benchmark	71
4.1. Generador de datos	71
4.1.1. Método de generación de grafos	71
4.1.2. Diseño	78
4.1.3. Implementación	81
4.2. Generador de operaciones	84
4.2.1. Operaciones de la carga de trabajo	85
4.2.2. Método de generación de datos de prueba	90
4.3. Evaluador de operaciones	91
4.3.1. Métricas y reportes	91
4.3.2. Diseño	92
4.3.3. Implementación	94
5. Evaluación del benchmark	98
5.1. Metodología de evaluación	98
5.2. Configuración y ejecución de los experimentos	99
5.3. Análisis de resultados	100
6. Conclusiones	112
Anexos	
A: Generador de datos	117
B: Generador y evaluador de operaciones	126

Bibliografia**149**

ÍNDICE DE FIGURAS

	página
2.1. Representación gráfica de un grafo.	18
2.2. Tipos de grafo	20
2.3. Ejemplo de una base de datos de grafos.	22
2.4. Ejemplo de una base de datos de grafos basado en Neo4j.	25
2.5. Modelo de datos propuesto por el generador de datos.	29
3.1. Representación gráfica de distintas distribuciones de nodos.	44
3.2. Distribución de aristas para un camino de longitud 1.	66
3.3. Distribución de aristas para un camino de longitud 2.	67
3.4. Distribución de aristas para un camino de largo K.	68
4.1. Modelo R-Mat	72
4.2. Ejemplo de funcionamiento de R-Mat, para grafos dirigidos y no dirigidos.	73
4.3. Almacenamiento de aristas en un arreglo, basado en el funcionamiento de R-Mat.	75
4.4. Representación gráfica del método de selección de aristas para 9 nodos.	77
4.5. Diagrama de flujo del generador de datos.	78
4.6. Diagrama de clases del generador de datos.	81
4.7. Diagrama de flujo del evaluador de operaciones.	97
5.1. Desviación estándar respecto a la consulta Q1.	104
5.2. Desviación estándar respecto a la consulta Q2.	105
5.3. Desviación estándar respecto a la consulta Q3.	105
5.4. Desviación estándar respecto a la consulta Q4.	106
5.5. Desviación estándar respecto a la consulta Q5.	106
5.6. Desviación estándar respecto a la consulta Q6.	107
5.7. Desviación estándar respecto a la consulta Q7.	107
5.8. Desviación estándar respecto a la consulta Q8.	108
5.9. Desviación estándar respecto a la consulta Q9.	108
5.10. Desviación estándar respecto consulta a la Q10.	109
5.11. Desviación estándar respecto consulta a la Q11.	109

5.12. Desviación estándar respecto consulta a la Q12.	110
5.13. Desviación estándar respecto consulta a la Q13.	110

ÍNDICE DE TABLAS

	página
3.1. Ejecución de la consulta de prueba 5	34
3.2. Resumen con los resultados de la ejecución del <i>benchmark</i> en la base de datos de grafos <i>DEX</i>	35
3.3. Resumen con los resultados de la ejecución del <i>benchmark</i> en la base de datos de grafos <i>Neo4J</i>	36
3.4. Ejecución de la consulta de prueba 5 en Dex.	36
3.5. Ejecución de la consulta de prueba 5 en Neo4J.	37
3.6. Ejecución de la consulta de prueba 6 en Dex.	37
3.7. Ejecución de la consulta de prueba 6 en Neo4j.	37
3.8. Ejecución de la consulta de prueba 7 en Dex.	38
3.9. Ejecución de la consulta de prueba 7 en Neo4j.	38
3.10. Ejecución de la consulta de prueba 8 en Dex.	39
3.11. Ejecución de la consulta de prueba 8 en Neo4j.	39
3.12. Ejecución de la consulta de prueba 9 en Dex.	39
3.13. Ejecución de la consulta de prueba 9 en Neo4j.	40
3.14. Ejecución de la consulta de prueba 10 en Dex.	40
3.15. Ejecución de la consulta de prueba 10 en Neo4j.	40
3.16. Ejecución de la consulta de prueba 11 en Dex.	41
3.17. Ejecución de la consulta de prueba 11 en Neo4j.	41
3.18. Resultados de aplicar las técnicas de selección de datos de prueba en Dex, con distribución de personas y páginas en base a la ley de potencias.	48
3.19. Resultados de aplicar las técnicas de selección de datos de prueba en Neo4j, con distribución de personas y páginas en base a la ley de potencias.	49
3.20. Salida de datos al aplicar las técnicas de selección de datos de prueba con una distribución de personas y páginas web basada en la ley de potencias.	50
3.21. Resultados de aplicar las técnicas de selección de datos de prueba en Dex, con distribución de personas en base a la ley de potencias y páginas en base a la distribución normal.	51

3.22. Resultados de aplicar las técnicas de selección de datos de prueba en Neo4j, con distribución de personas en base a la ley de potencias y páginas web en base a la distribución normal.	52
3.23. Salida de datos al aplicar las técnicas de selección de datos de prueba con una distribución de personas basada en la ley de potencias y páginas web en base a la distribución normal.	53
3.24. Resultados de aplicar las técnicas de selección de datos de prueba en Dex, con distribución de personas en base a la distribución normal y páginas web distribuidas en base a la ley de potencias.	54
3.25. Resultados de aplicar las técnicas de selección de datos de prueba en Neo4j, con distribución de personas en base a la distribución normal y páginas web distribuidas en base a la ley de potencias.	55
3.26. Salida de datos al aplicar las técnicas de selección de datos de prueba con una distribución de personas basada en la distribución normal y páginas web distribuidas en base a ley de potencias.	56
3.27. Resultados de aplicar las técnicas de selección de datos de prueba en Dex, con distribución de personas y páginas en base a la distribución normal.	57
3.28. Resultados de aplicar las técnicas de selección de datos de prueba en Neo4j, con distribución de personas y páginas en base a la distribución normal.	58
3.29. Salida de datos al aplicar las técnicas de selección de datos de prueba con una distribución de personas y páginas web basada en la distribución normal.	59
3.30. Comparación de grados de nodos en una distribución de ley de potencias.	60
3.31. Comparación de grados de nodos en una distribución normal.	60
3.32. Comparación de técnicas respecto a su tiempo de ejecución promedio.	62
3.33. Comparación de técnicas respecto a su salida de datos promedio.	63
3.34. Promedio de resultados entre salida y tiempo de ejecución.	64
3.35. Tiempo de ejecución de la consulta “Obtener la cantidad de amigos de los amigos de los amigos de P”.	69
3.36. Resultados de ejecución de la consulta “Obtener la cantidad de amigos de los amigos de los amigos de P”.	70

- 5.1. Tiempos obtenidos en Dex al utilizar datos de entrada contenidos en la mediana. 101
- 5.2. Tiempos obtenidos en Neo4j al utilizar datos de entrada contenidos en la mediana. 102
- 5.3. Tiempos obtenidos en Dex al utilizar datos de entrada aleatorios. . . 103
- 5.4. Tiempos obtenidos en Neo4j al utilizar datos de entrada aleatorios. . 104