

## TABLA DE CONTENIDOS

	página
<b>Dedicatoria</b>	<b>I</b>
<b>Agradecimientos</b>	<b>II</b>
<b>Tabla de Contenidos</b>	<b>III</b>
<b>Indice de Figuras</b>	<b>VI</b>
<b>Indice de Algoritmos</b>	<b>VIII</b>
<b>Resumen</b>	<b>IX</b>
<b>Abstract</b>	<b>x</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Descripción del problema . . . . .	1
1.2. Hipótesis . . . . .	2
1.3. Objetivos . . . . .	3
1.3.1. Objetivo general . . . . .	3
1.3.2. Objetivos específicos . . . . .	3
<b>2. Antecedentes</b>	<b>4</b>
2.1. Introducción a grafos . . . . .	4
2.2. Formatos de codificación de grafos . . . . .	7
2.2.1. Trivial Graph Format (TGF) . . . . .	7
2.2.2. DOT Language . . . . .	7
2.2.3. GraphML . . . . .	9
2.3. Bases de datos de grafos . . . . .	10
2.3.1. DEX . . . . .	10
2.3.2. Neo4j . . . . .	11
2.3.3. FlockDB . . . . .	11
2.3.4. HyperGraphDB . . . . .	12
2.3.5. InfoGrid . . . . .	12

2.3.6.	Otras bases de datos de grafos . . . . .	13
2.3.7.	Lenguajes de consulta para bases de datos de grafos . . . . .	13
2.4.	Benchmarks . . . . .	14
2.4.1.	Clasificación de los benchmarks . . . . .	16
2.4.2.	TPC . . . . .	16
2.4.3.	HPC Graph Analysis . . . . .	18
2.4.4.	RBench: The RDF Benchmark . . . . .	19
2.4.5.	Conclusiones sobre los benchmarks . . . . .	20
<b>3.</b>	<b>Trabajo preliminar</b>	<b>21</b>
3.1.	Análisis y especificación de requerimientos del benchmark . . . . .	21
3.2.	Caso de estudio: datos de redes sociales . . . . .	24
3.2.1.	Redes sociales orientadas a relaciones de amistad . . . . .	25
3.2.2.	Redes sociales no orientadas a relaciones de amistad . . . . .	26
3.2.3.	Esquema de datos para una red social . . . . .	28
<b>4.</b>	<b>Métodos de generación de grafos</b>	<b>31</b>
4.1.	Generación de datos reales usando Hi5 . . . . .	32
4.1.1.	Generador de grafo por amplitud . . . . .	41
4.1.2.	Generador de grafo basado en relaciones de importancia . . . . .	42
4.1.3.	Generador de grafo por poda natural . . . . .	44
4.1.4.	Generador de grafo por poda natural optimizada . . . . .	45
4.1.5.	Comparación de métodos de generación . . . . .	47
4.2.	Generación de grafos sintéticos . . . . .	48
4.2.1.	R-MAT: Matriz recursiva . . . . .	49
4.2.2.	R-MAT*: Matriz recursiva adaptada a lista de adyacencia . . . . .	52
4.2.3.	SNG: <i>Social Network Generator</i> . . . . .	53
4.2.4.	SNG*: <i>Social Network Generator</i> simplificado . . . . .	57
4.2.5.	R-LIST: Lista recursiva . . . . .	59
4.2.6.	R-LIST*: Lista recursiva limitada . . . . .	62
4.2.7.	R-LIST+: Lista recursiva con estructura superior . . . . .	64
4.2.8.	Comparación de métodos de generación . . . . .	68
<b>5.</b>	<b>Gestor de pruebas básico</b>	<b>74</b>
5.1.	Consultas . . . . .	74

5.1.1. Conjunto de consultas a evaluar . . . . .	75
5.2. Gestión de pruebas . . . . .	79
5.3. Caso de Prueba: DEX . . . . .	82
<b>6. Conclusiones</b>	<b>85</b>
6.1. Trabajos futuros . . . . .	87
<b>Bibliografía</b>	<b>88</b>

## INDICE DE FIGURAS

	página
2.1. <i>Elementos de un grafo</i> . . . . .	4
2.2. <i>Aplicación real: Red de flujo</i> . . . . .	6
2.3. <i>Aplicación real: Mapa político</i> . . . . .	6
3.1. <i>Estructura de Red Social</i> . . . . .	24
3.2. <i>Estructura de Facebook</i> . . . . .	26
3.3. <i>Estructura de Google+</i> . . . . .	26
3.4. <i>Estructura de Hi5</i> . . . . .	27
3.5. <i>Estructura de Twitter</i> . . . . .	28
3.6. <i>Esquema de la red social</i> . . . . .	30
4.1. <i>Diagrama de Paquetes de Crawler Hi5</i> . . . . .	33
4.2. <i>Diagrama de entidad relación de base de datos en bruto</i> . . . . .	35
4.3. <i>Diagrama de entidad relación de base de datos con formato de grafo</i> . . . . .	36
4.4. <i>Diagrama de clases de sistema extracción y generación de datos Hi5</i> . . . . .	39
4.5. <i>Grafo Original</i> . . . . .	41
4.6. <i>Grafo generado por amplitud</i> . . . . .	41
4.7. <i>Grafo generado en base a relaciones de importancia</i> . . . . .	43
4.8. <i>Grafo generado por poda natural</i> . . . . .	45
4.9. <i>Grafo generado en base a poda natural optimizada</i> . . . . .	46
4.10. <i>Modelo R-MAT</i> . . . . .	50
4.11. <i>Distribución R-MAT</i> . . . . .	51
4.12. <i>Representación de matriz de adyacencia en lista de adyacencia</i> . . . . .	53
4.13. <i>Distribución R-MAT*</i> . . . . .	54
4.14. <i>Distribución SNG</i> . . . . .	56
4.15. <i>Anomalía en distribución de ley de potencias</i> . . . . .	58
4.16. <i>Distribución SNG*</i> . . . . .	58
4.17. <i>Modelo R-LIST</i> . . . . .	60
4.18. <i>Distribución R-LIST</i> . . . . .	62
4.19. <i>Límite a profundidad</i> . . . . .	63
4.20. <i>Distribución R-LIST*</i> . . . . .	64

4.21. <i>Modelo R-LIST+</i> . . . . .	65
4.22. <i>Ejemplo R-LIST+</i> . . . . .	67
4.23. <i>Comparación de generadores de grafos sintéticos con escala 10</i> . . . . .	69
4.24. <i>Comparación de generadores de grafos sintéticos con escala 20</i> . . . . .	70
4.25. <i>Comparación de generadores de grafos sintéticos con escala 26</i> . . . . .	71
4.26. <i>Comparación de tiempos de generación de grafos sintéticos</i> . . . . .	72
5.1. <i>Consulta 1</i> . . . . .	75
5.2. <i>Consulta 2</i> . . . . .	76
5.3. <i>Consulta 3</i> . . . . .	76
5.4. <i>Consulta 4</i> . . . . .	76
5.5. <i>Consulta 5</i> . . . . .	77
5.6. <i>Consulta 6</i> . . . . .	77
5.7. <i>Consulta 7</i> . . . . .	77
5.8. <i>Consulta 8</i> . . . . .	77
5.9. <i>Consulta 9</i> . . . . .	78
5.10. <i>Consulta 10</i> . . . . .	78
5.11. <i>Consulta 11</i> . . . . .	78
5.12. <i>Gestión de pruebas</i> . . . . .	81

## INDICE DE ALGORITMOS

	página
4.1. <i>Generador de grafo por amplitud</i> . . . . .	42
4.2. <i>Generación de grafo en base a relaciones de importancia</i> . . . . .	44
4.3. <i>Generación de grafo por poda natural</i> . . . . .	45
4.4. <i>Generación de grafo por poda natural optimizada</i> . . . . .	47
4.5. <i>R-MAT</i> . . . . .	51
4.6. <i>R-MAT*</i> . . . . .	54
4.7. <i>SNG</i> . . . . .	55
4.8. <i>SNG*</i> . . . . .	59
4.9. <i>R-LIST</i> . . . . .	61
4.10. <i>R-LIST*</i> . . . . .	63
4.11. <i>R-LIST+</i> . . . . .	66