



ESCUELA
NACIONAL
DE ESTUDIOS
SUPERIORES

UNIDAD MORELIA

Protocolo de Investigación

Estudio y Modelación de Grafos de Vecindad Relativa (RNG)

Universidad Nacional Autónoma de México

ENES Morelia

28/04/2024

Asesor: Daniele Colosi

Alumno: Miguel Ángel Zamorano Presa

Licenciatura en Tecnologías de la Información para las Ciencias

Contents

0.1	Antecedentes	2
0.2	Introducción	3
0.3	Objetivos	3
0.3.1	Objetivos generales	3
0.3.2	Objetivos específicos	4
0.4	Metodología	5
0.5	Resultados Preliminares	5
0.5.1	Modelado	7
0.5.2	Próximas líneas de investigación	8
	Bibliografía	9

0.1 Antecedentes

Los grafos son un instrumento matemático compuesto de dos objetos(puntos y aristas) utilizado para describir y modelar todo tipo de sistemas complejos.

Los grafos de vecindad relativa (RNG, Relative Neighborhood Graphs) fueron introducidos por Toussaint [1980], son un tipo de grafos derivados de los llamados grafos de proximidad.

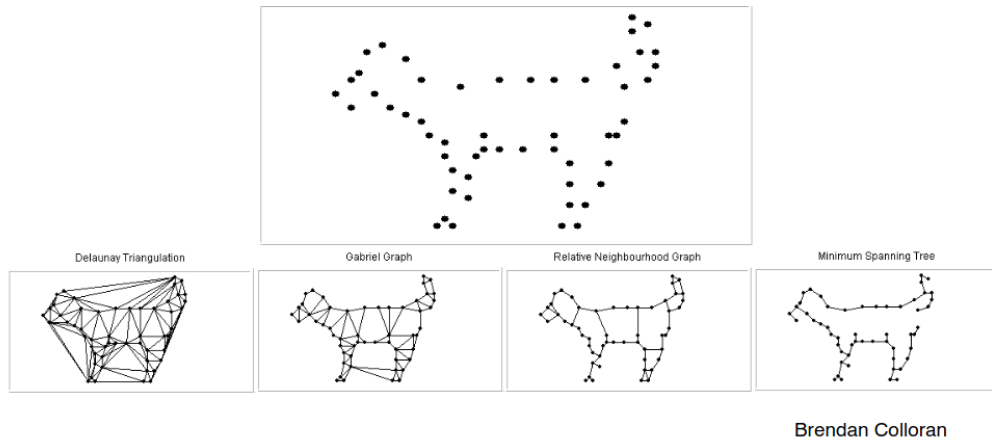
Estos grafos se generan a partir de un conjunto de puntos, en donde dos nodos están conectados con base a el siguiente criterio geométrico:

Se definen en término de una relación geométrica entre los puntos: una arista es compartida por dos puntos si y solo si se cumple que la distancia entre u y v es menor o igual a la distancia máxima entre cada uno de estos puntos y cualquier otro punto z del grafo:

$$d_{uv} \leq \max\{d_{uz}, d_{vz}\}$$



Other Proximity Graphs: Relative Neighbourhood and Gabriel graphs



La figura es un gran apoyo visual para distinguir y comparar los diferentes tipos de patrones generados por los principales grafos de proximidad aplicados al mismo conjunto de puntos.

Toussaint [1980] introduce los RNG como una alternativa a los grafos existentes para extraer patrones, el demuestra que las RNGs son en realidad un derivado de los DT(Delaunay triangulation) y y un supraconjunto de los MST(Minimal spanning tree), también nos presenta 2 algoritmos para generar los RNG.

El primero y el original,de complejidad $O(n^3)$ que es el que utilizamos y un segundo donde utilizando la propiedad que las RNGs son un subconjunto de los Diagramas de Voronoi,primero calculan dicho grafo y después le aplican una limpia de aristas de complejidad $O(n^2)$.

En el artículo de Toussaint [1980] habla sobre los problemas abiertos y su aplicabilidad en:

1. El problema de la extracción de estructuras geométricas a partir de un conjunto finito de puntos embebidos a un plano.
2. El potencial del espectro de posibles problemas abiertos relacionados a probabilidad geométrica, este último fue fuente de inspiración para esta investigación.

Otras áreas de aplicación fueron surgiendo como el caso de Foster et al. [2022] donde extiende el problema abierto publicado por Toussaint [1980] sobre la construcción de RNGs en espacios de alta dimensionalidad y por lo tanto la capacidad de reconocer vecinos relacionados en un orden mas alto, si bien no es tema de nuestro interés por el momento es interesante conocer la diversidad de estudios al rededor de los trabajos de Toussaint [1980].

Las RNG han encontrado aplicación en diversos campos: Modelado 3D, renderización Toussaint [1980], en técnicas para minería de datos Zemel and Carreira-Perpiñán [2004], por mencionar algunos.

0.2 Introducción

La literatura existente de las RNGs se limita al estudio de grafos embebidos a planos y no a RNGs que viven en planos [superficies] con curvaturas.

En este trabajo, se extienden la investigación de las propiedades estadísticas de las RNG en superficie plana [se investigan las propiedades estadísticas de las RNG] a una superficie esférica unitaria (distribución de las aristas, grado promedio, correlaciones, etc.) y se proponen modelos de grafos random que aproximen los grafos RNG, con el objetivo de reducir la complejidad de construcción de los grafos RNG a un generador de RNGs compuesto de únicamente dos parámetros (el número de puntos y el grado promedio [la distancia entre nodos]), en lugar de tener que hacer la evaluación n cuadrática de tener que calcular la matriz de distancias entre los puntos y estar evaluando el criterio geométrico.

Por qué esferas?

Las tomamos a conveniencia por varias razones:

1. En la literatura solo se han estudiado RNG en planos sin curvaturas.
2. Los modelos a una escala de esfera unitaria es una herramienta conveniente para estudiar y conducir experimentos de manera sencilla y las usaremos para modelar grafos que embebidos a una esfera y extraer propiedades.
3. La geometría de la superficie de una esfera es conocida e históricamente estudiada, existe una parte de técnicas de la geometría analítica euclidiana tradicional que tiene sus formulas equivalentes a este espacio y geometría esférica.
4. Varias cosas interesantes como el estudio planetas [redes distribuidas sobre la superficie del planeta], inclusive el cerebro y el modelado de huellas dactilares tienen una geometría aproximadamente esférica elipsoidal.

0.3 Objetivos

0.3.1 Objetivos generales

Estudio estadístico descriptivo de los grafos de vecindad relativa (RNG, Relative Neighborhood Graphs) sobre la superficie de la esfera unitaria y proponer modelos aproximados de RNG por medio de gráficas random y de esa forma reducir la complejidad computacional .

0.3.2 Objetivos específicos

- 1.1. Extraer las propiedades estadísticas relevantes de las gráficas RNG.
- 1.2. Relacionar las propiedades de las RNG con la superficie del grafo.
- 1.3. Implementar una definición de entropía como área dentro de la cual puede moverse los vértices de las gráficas RNG sin afectar su matriz de adyacencia.
2. Proponer un modelo que genere grafos random que aproxime una RNG utilizando nuestros resultados del análisis estadístico en forma de heurísticas (distribución de número de links y distribución del valor de distancia de links).

Limitaciones de las RNG

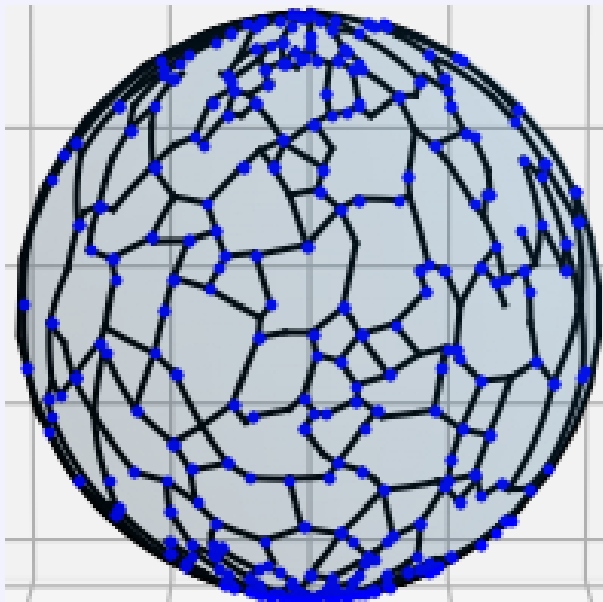
- Identificar una RNG es computacionalmente demandante
- En la literatura solo se han estudiado RNG en planos sin curvaturas

Hipótesis de trabajo

- Es posible reducir la complejidad de las RNG de manera aproximada

$$\text{no RNG} \simeq \text{RNG}$$

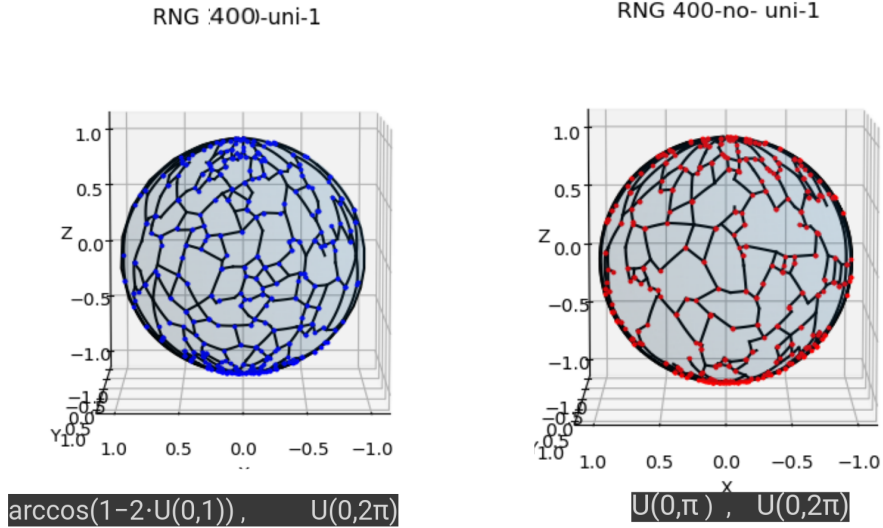
- Es posible extender las RNG a superficies con curvatura



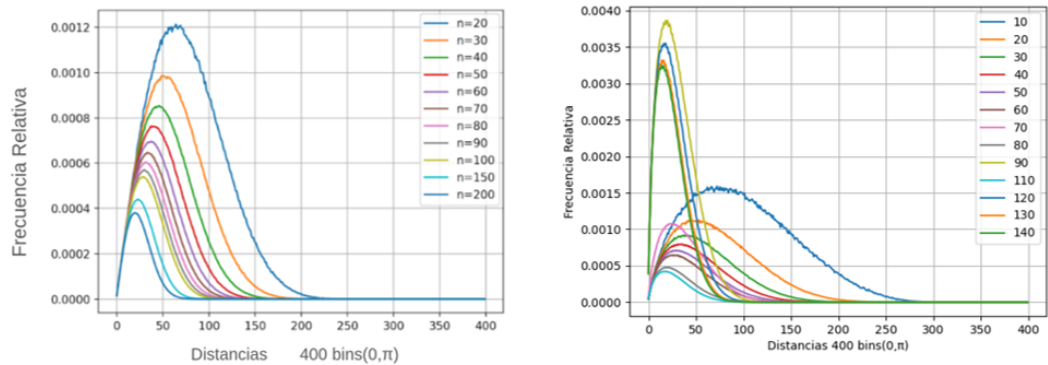
0.4 Metodología

Simulaciones y estudio estadístico

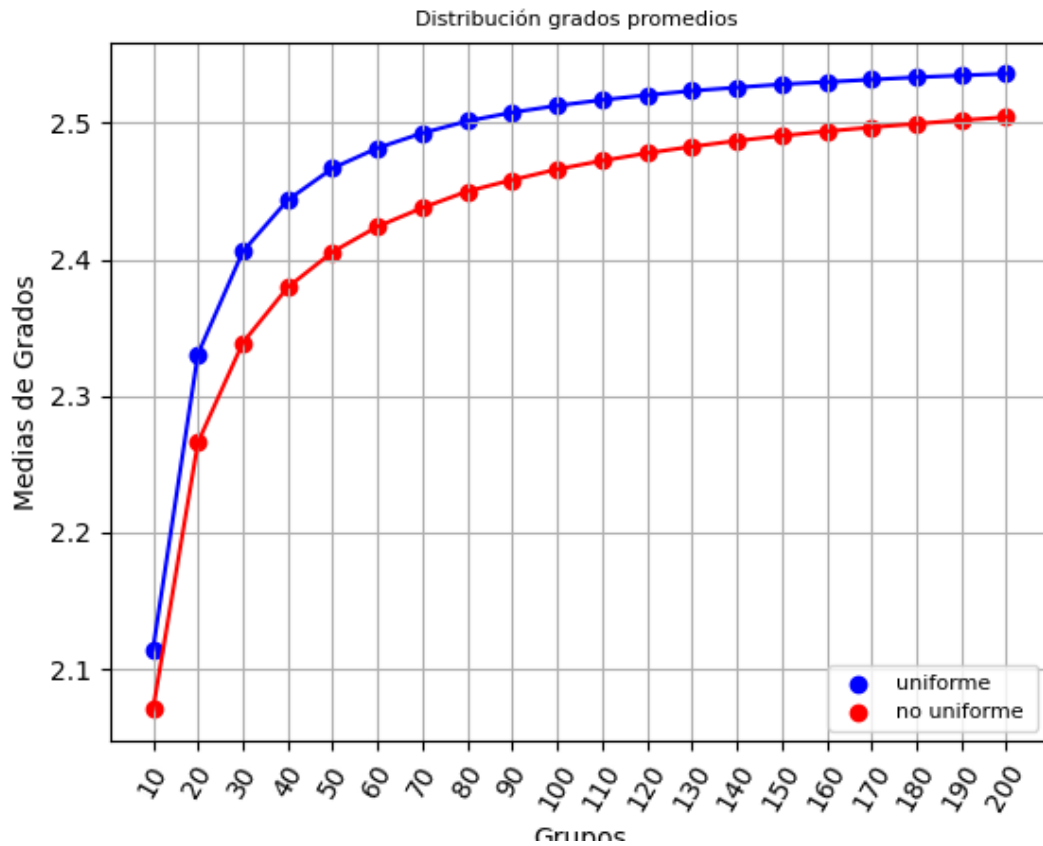
1. Se crean dos categorías de puntos en la superficie esférica: 1) puntos distribuidos uniformemente en la esfera: 2) puntos no distribuidos uniformemente que terminaba por concentrar los puntos en los polos. Esto con el fin de recabar más informaciones estadísticas sobre las propiedades de las RNGs.
2. A cada conjunto de puntos en la esfera está asociada unívocamente una RNG. Se generan 10^5 para cada grupo de n puntos con $n \in \{10, 20, \dots, 200\}$. Más allá de 200 es computacionalmente muy demandante.
3. Se estudian las propiedades estadísticas: distribución de las distancias de las aristas de las RNGs (moda, media y mediana); grado promedio de los nodos
4. Se representan con leyes generales.
5. Se buscan correlaciones entre las cantidades medidas
6. Construimos un visualizador 3d, en la figura mostramos dos grafos generados con mismo número de puntos pero diferente distribución, la azul la uniforme y la roja la no uniforme.



0.5 Resultados Preliminares

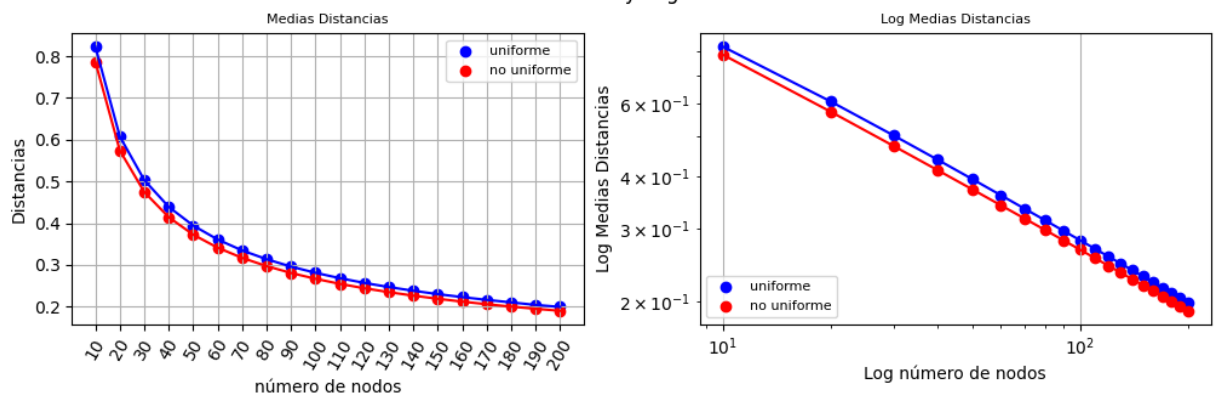


Las dos gráficas de arriba muestran la distribución de las distancias de cada uno de los grupos para cada tipo de distribución, la uniforme y no uniforme.

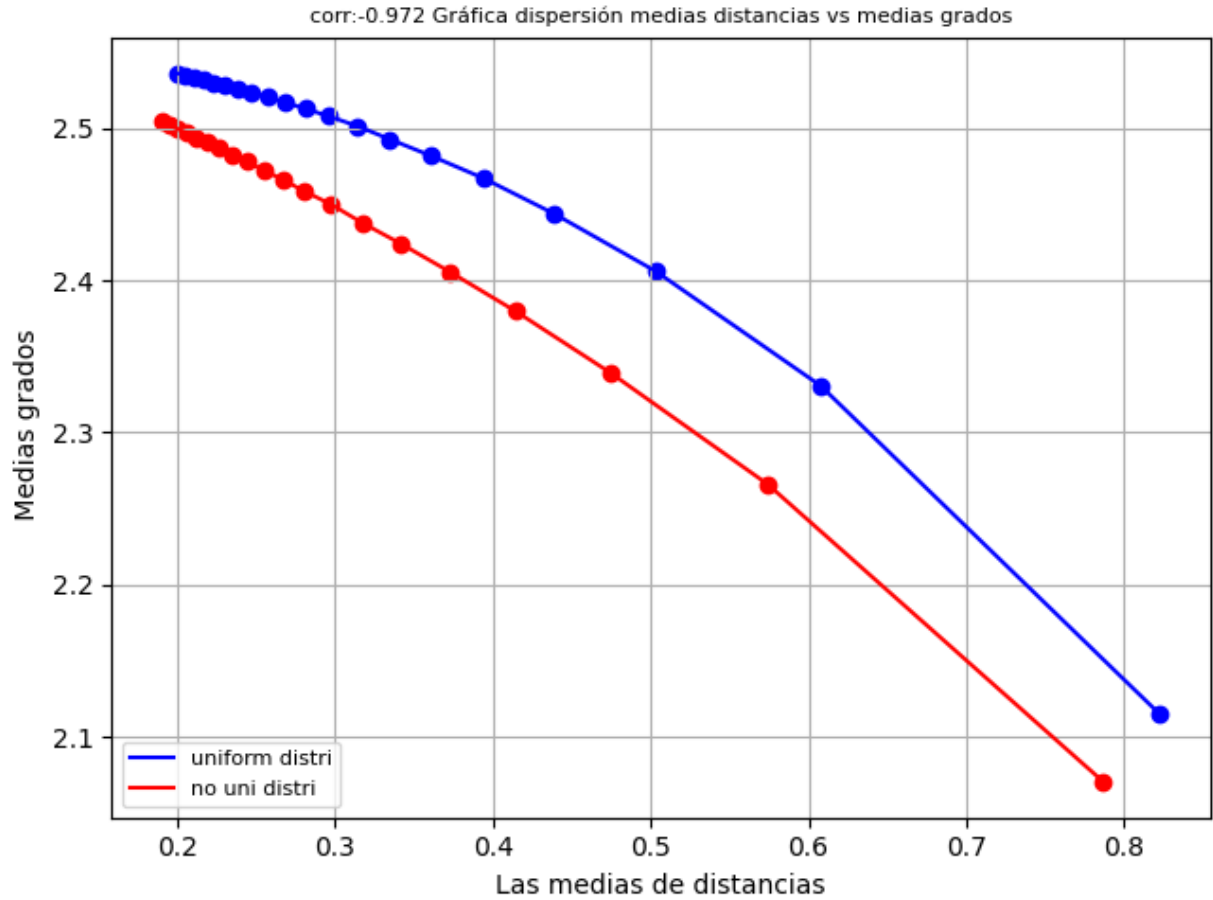


Sin importar la distribución fuera uniforme o no uniforme observamos que al aumentar el número de puntos el grado promedio converge a 2.5. [Al revés hay que subrayar que el grado promedio tiene un límite para $n \rightarrow \infty$ diferente para las dos distribuciones.]

Gráficas Medias Distancias y Log Medias Distancias



Las distribuciones obtenidas siguen una ley de potencia.



	Uniforme	No Uniforme
Correlación de Pearson con grado promedio		
Media	-0.972000	-0.992000
Moda	-0.963000	-0.993000
Mediana	-0.975000	-0.995000

Los resultados de las correlaciones son interesantes porque nos indica que la propiedad de los grafos RNG que la relación de la distancia con respecto al número de grados es inversamente proporcional.

0.5.1 Modelado

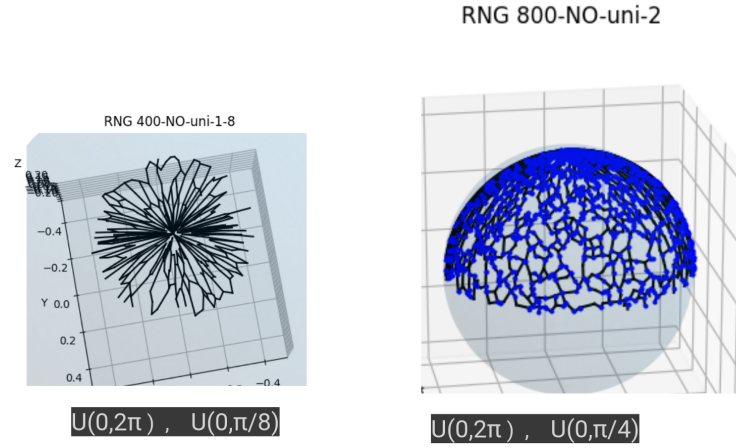
Usando las distribuciones encontradas en la fase 1 para construir una función de probabilidad, esta función asignará a cada par de puntos la probabilidad de que dos puntos tengan o no link. Se generarán las correspondientes gráficas random y se medirá el error con respecto a la RNG (para el mismo conjunto de puntos). Se utilizarán 4 medidas de error:

1. mean square error(MSE)
2. mean absolute percentage error (MAPE),
3. the relative entropy (RE)
4. the cosine proximity (CP).

Los resultados de nuestras funciones candidatas deberán minimizar el error de bondad entre los resultados del algoritmo probabilístico con respecto a los resultados del algoritmo original por un margen significativo.

0.5.2 Próximas líneas de investigación

Analizar y entender el papel que tiene la curvatura de la esfera sobre nuestros resultados.



Se reproducen las simulaciones realizadas restringiendo el espacio de distribución de puntos de manera gradual como se muestra en la imagen, hasta quedarnos solo con caparazones de longitudes reducidas.

El objetivo es proponer una nueva definición de entropía, la media de las áreas dentro de la cual cada punto puede moverse los vértices de gráficas RNG sin afectar su matriz de adyacencia, es decir, la región de espacio donde cada uno de los puntos podrían ubicarse sin modificar la configuración de la RNG original.

Utilizando el criterio de entropía que obtengamos podremos proponer un modelo estadístico mecánico que logre describir las propiedades macroscópicas de las RNG a partir del criterio de entropía.

Bibliography

- D. Colosi and P. Dall'Olio. Relative neighbourhood graphs on the sphere. 2021.
- C. A. Foster, B. Sevilmis, and B. B. Kimia. Generalized relative neighborhood graph (grng) for similarity search. *ArXiv*, abs/2208.10022, 2022. URL <https://api.semanticscholar.org/CorpusID:251719309>.
- G. T. Toussaint. The relative neighbourhood graph of a finite planar set. *Pattern recognition*, 12(4): 261–268, 1980.
- R. Zemel and M. Carreira-Perpiñán. Proximity graphs for clustering and manifold learning. In L. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems*, volume 17. MIT Press, 2004. URL https://proceedings.neurips.cc/paper_files/paper/2004/file/dcda54e29207294d8e7e1b537338b1c0-Paper.pdf.