**Khanam - The homophily principle in social network analysis: A survey (2022)**

1. Idea principal.

Review de la homofília en redes sociales. Principalmente enfocado en trabajos relacionados con redes levantadas con datos de redes sociales (N~10^4). Valorable del marco:

Homophily was first divided into two types: *status*

homophily and *value* homophily [56, 60, 110]. The concept of status homophily is centered

on an individual’s social standing, implying that people in similar social situations are more

likely to mix. Status homophily also includes the major sociodemographic dimensions such

as race, ethnicity, sex, or age, and procured characteristics like religion, education, occupation,

or behaviour patterns [4]. Value homophily, on the other hand, is based on the similarity

of people’s thinking, leading to the assumption that people with similar thoughts are more

likely to interact with others, even though differences may lie in their social positions. It

includes the internal states that are thought to shape an individual’s orientation towards the

future [25, 27, 79]. Homophily is a broad field with also some strong connections to social

cognitive mechanisms [30]. Albert Bandura proposed the Social Cognitive Theory, which

stresses the dynamic interplay between individuals (personal variables), their behaviour,

and their settings [3]. The behaviour between the individuals includes the heartfelt emotions

(empathy) that the individuals feel for one another.

1. Métodos.

Buscaron [*“social media”*] *AND [“Homophily” OR “degree distribution”]* en google scholar. Filtraron los de h-index sobre 50, o aquellos de journals de Q1 o Q2, o aquellos con más de 100 citas.

1. ¿Cómo se relaciona con el tema de investigación?

De gran ayuda en caso de centrarse el estudio en redes grandes.

**Currarini - A simple model of homophily in social networks (2016)**

1. Idea principal.

Se ha mostrado que la homofília observada en redes tiene que ver con las oportunidades de encuentro que en principio están sesgadas. Este paper presenta se estudia un modelo de los micro-fundamentos de las búsquedas de encuentro que generan los encuentros sesgados, y que sus predicciones de equilibrio son consistentes con las relaciones no-lineales (en relación al índice de Coleman) observadas en redes de nominación de amistad y en los matrimonios de EEUU. El modelo es un juego de costos y beneficios. La no linealidad se refiere a la forma de joroba del índice de Coleman con respecto a la proporción de la población que está en ese grupo. Se permite la formación de redes con información fluyendo en una dirección (como seguir páginas de twitter que no requiere confirmación de la otra parte), como de dos direcciones (matrimonios).

1. ¿Cómo se relaciona con tema de investigación?

Presenta un modelo para explicar el sesgo en las oportunidades de meeting, que resultan en una homofília observada en la red. No es adecuado para gerenar una red.

**Lee - Homophily and minority-group size explain perception biases in social networks (2019)**

1. Idea principal.

El paper intenta explicar el sesgo de percepción de grupos minoritarios a partir de una mirada de teoría de redes. En específico, investiga el impacto de distintas propiedades de redes en el sesgo de percepción. El modelo acepta **asimetría de homofília**: el hecho que un grupo A sienta más cercanía a otro grupo B que el que siente B hacia A. Un factor importante es el tamaño ralativo entre los grupos.

1. Metodología.

Conducen tres **encuestas** con n~100 para calcular homofília en la red personal. En su notación, se entiende h \in [0,1/2] == heterofília, y se entiende h \in [1/2, 1] == homofília. Además, crean una manera de producir redes **scale-free con homofília** para tener resultados numéricos. Por último, solucionan el set de ecuaciones para presentar una solución **analítica** y así contrastarlo con las otras entradas.

1. ¿Cómo se relaciona a tema de investigación?

Usaron una manera de crear una red scale-free con homofília, y tiene github. El problema es que se entiende la homofília como se entiende el *inbreed homophily*: hay dos grupos, estás dentro o estás fuera. Así, h es una medida de cercanía de un grupo con otro. Es solo un número, aunque se podría pensar en 4: h\_mm, h\_MM, h\_mM, h\_Mm; donde m==minoría, M==mayoría.

Presenta un problema si lo que quiero es fijar una red (no cambiarla cada iteración) donde hayan varios valores de h, para que h\_windows no sea trivial, porque básicamente hay 4 valores globales de h. Es decir, la noción h no está dentro de cada par de nodos, sino que es un valor global.

**Carrington - Log-linear distance models of homophily in small groups (2016)**

1. Idea principal.

El artículo se centra en demostrar las ventajas de usar el ‘log-linear distance model’ en la evaluación de la homofília en un conjunto de grupos pequeños. En específico, hace un review de su uso en la homofília de edad y género mostrada por grupos de ‘co-offenders’ delictuales. Además, una suerte de tutorial de cómo aplicar ese método de homofília de una o más variables en cualquier grupo pequeño.

1. Metodología.

Así como en el estudio de parezas sexuales por raza, se comienza preguntando a un conjunto de individuos: ¿cuántos encuentros sexuales has tenido con personas de esta raza? Así, se forma un matriz ‘squeare cross-tabulation’ H(i,j). El modelo logarítmitco tiene un fator de intercepto, un factor para las columnas, otro para las filas, y un cuarto factor que permite la diferencia \beta|i-j|, de donde se puede calcular la *distancia social* \beta entre las categorías i y j. No todo par de categoría merece una distancia social, ya que algunos bastan con explicarse con homofília.

Este modelo, a diferencia del índice de coleman, permite el uso de **variables ordinales** (cualitativas con orden o jerarquía), y aplica a **interacciones diádicas**: con qué otra persona discutes cosas importantes, no con qué grupo lo haces. Se presenta una **extensión la grupos pequeños**.

This phenomenon

is limited to ‘small’ groups because in larger groups it is

unrealistic to assume that all members interact with all other

members (Mayhew et al., 1995: 25).

1. ¿Cómo se relaciona con tema de investigación?

Además de la potencial ayuda en aclarar cómo usar una forma más moderna de medir homofília en redes, este paper es (o fue) crucial pues permite desarrollar un concepto de ‘sesgo de selección de pares’ de una forma contínua. Esto lo aclara al criticar al índice de Coleman por presentar una conceptualización binaria de ‘same/different’ y no permitir una noción de *distancia social*, además de no permitir inferencia estadística. Otras ayudas se brindan:

Homophily has been defined as ‘the principle that a contact

between similar people occurs at a higher rate than among

dissimilar people’ (McPherson et al., 2001: 416). Mayhew et

al. (1995) characterise homophily as ‘the social distance

interaction hypothesis’: ‘The frequency with which people

enter into face-to-face association (contact, interaction) is

inversely related to the social distance between them’ (p. 20)….

Explanations of observed homophily generally fall into three categories:

1. *Population composition*: This might be characterised as the‘null’hypothesis, baseline homophily.
2. *Preference*: For various reasons, people *prefer* to associate and collaborate with those who are similar to them.
3. *Social structure*: Features of the social structure apart from overall population composition may constitute‘opportunity structures’ that favour homophilous contact and collaboration. Algo como ***Currarini (2016)*.**

Marsden 1988 aplica este *log-lineal distance model* para analizar la homofília en redes personales, entre quienes ‘discuten cosas importantes’, construyedo *cross-tabulations* entre edad, educación, raza, religión y género.

**Iijima - Social distance and network structures (2017)**

1. Idea principal.

Las redes sociales observadas por lo general son de un tipo específico de ‘small-world’.

*“Real-world social networks often exhibit a set of characteristic structural properties that do not occur jointly as often in different kinds of complex networks (i.e. technological, informational, biological).* *Real-world social networks often exhibit* ***high levels of clustering****,* ***positive degree assortativity****,* ***short average path lengths*** *(small-world property) and* ***right-skewed*** *but rarely power law degree distributions”. TALAGA (2020)*

Iijima propone analizar a qué punto esta topología de red puede ser relacionado con agentes midiendo distancias sociales.

Hace la diferencia que en las redes de e-mail no se necesitan aspectos similares entre el remitente y dirigido para que este lo reciba, pero en las redes de coautoría ambas partes tienen más intereses similares. Ambas redes tienen distinto avg. path length y clusterización. ¿Se puede explicar por lo similares que sean sus intereses?

1. Motodología.

Proponen el constructo ***type space***, que al parecer no está basado en literatura provia, de tal manera que cada agente está equipado con un vector m-dimensional cuyas componentes indican alguna componente carácterística del agente, como su lugar de trabajo, locations, tastes, political views, and so forth. Así, definen ***social distance*** como las distancias en este *type space*, y representan la cantidad de obstaculos a las relaciones de los agentes; aplicando un principio homofílico para decir que se formarán links entre aquellos que tengan una baja *social distance* entre ellos, así como lo hizo Akerlof (1997) to study social decisions such as choices of educational attainment and childbearing.

Postulan un modelo *Utility-based* de formación de redes (aunque no escrita en un lenguaje de programación) basado en los costos y beneficios de la formación de link (*That is, agents obtain higher benefits from linking to a closer agent in the type space, while needing to pay a fixed cost to maintain a link*), llamado ***cutoff rule model*** en el que los agentes forman links si la (Kth smaller distance?) es menor que un cutoff dado exógenamente. Siempre que se viole la desigualdad triangular (k<m), se predice la propiedad ‘small-world’. Este requierimiento se ha observado en cognitive psychology.

*Modeling heterogeneity* One of the standard assumptions often

used in the economics literature is that people are partitioned into several groups,

and the relationships within a group cost less than the relationships across groups (see

Currarini et al. 2009 and Jackson and Rogers 2005). Such a modeling assumption leads to

networks with low average path lengths and high clustering coefficients, well observed

properties of real world networks. While the partitioning of agents implies that an agent

belongs to exactly one group, in reality an agent may well belong to multiple groups

(e.g., workplaces or local communities) or, more generally, she might be associated with

attitudes or tastes on a variety of aspects (e.g., political/ethical views or tastes in music/

sports)

*Homophily*. We assume that social distance describes the similarity between agents’

characteristics, and that similar agents are connected to each other. This assumption

is motivated by a well observed sociopsychological tendency called homophily (see Pin

and Rogers 2016 for a recent survey on this topic). Our kth norm provides a tractable

parameterization of homophily and enables us to analyze how different types of homophily

result in different network structures

…the interpretation of the kth norm is that if k is large, agents

care about many aspects of other’s types, while if k is small, then they care about very

few aspects of others’ types.

Cada agente tiene su propio cutoff de distancia social. Es decir, si la distancia social es menor que el menor cutoff de los nodos i,j para una cierta medida k-distancia, entonces hay nodo. Esta forma de armar la red es independiente de los valores de m y k.

1. ¿Cómo se relaciona con mi tema de investigación?

Usa el *type space* y la *social distance*. Pero además de eso, **explica cómo surgen las redes small-world desde un utility-based formation machanism.**

Ahora bien, un problema es que no se puede determinar m para cada situación en la que no sea seteada externamente.

**Talaga - Homophily as a process generating social networks: insights from SDA model (2020)**

1. Idea principal.

*“Real-world social networks often exhibit a set of characteristic structural properties that do not occur jointly as often in different kinds of complex networks (i.e. technological, informational, biological).* *Real-world social networks often exhibit* ***high levels of clustering****,* ***positive degree assortativity****,* ***short average path lengths*** *(small-world property) and* ***right-skewed*** *but rarely power law degree distributions”.*

*“well-known generic models such as Erdos-Rényi random graphs (Erdos & Rényi 1959) or preferential attachment networks (Barabasi & Albert 1999) that may not fit well to the social problem at hand.”*

La motivación es similar a la Iijima: *we examine the extent to which homophily is sufficient to produce the typical structural properties of social networks.*

Pero acá no usan el modelo basado en utilidad, sino que usan el **Social Distance Attachment (SDA) model** (Boguñá et al. 2004), un tipo particular de Random Geometric Graph (RGG) (Dall & Christensen 2002; Krioukov 2016).

Argumentan que la construcción de RGG se corresponde directamente al principio de homofilia, por lo que concluyen que la homofília, combinado con un random edge rewiring, es suficiente para reproducir muchos de las características típicas de las redes sociales reales.

1. Metodología

La **estructura social** es la distribución de posiciones sociales (coordenadas del vertor en ***type space***, aquí llamado al parecer ***social spaces***: *in which actors are represented as points and relations between them follow from distances in the space (Blau 1977; McPherson 2004; Bourdieu 1986, 1989).*) y las posiciones relativas determinan la probabilidad de ocurrencia de un lazo de un cierto tipo entre dos agentes.

*It is important to note that there are several different kinds of homophily. Usually in the literature one can find a distinction between value and status or choice and induced homophily (McPherson & Smith-Lovin 1987; McPherson et al. 2001); Value/choice homophily refers to similarity between agents that is due entirely to their more or less conscious preferences; On the other hand status/induced homophily refers to similarity that follows from structural constraints imposed on social agents such as geographical distance, racial and ethnic segregation etc.*

*…distinction between value/choice and status/induced homophily also reduces just to the issue of interpretation of dimensions of a social space.*

El modelo de formación de lazos requiere de una distancia característica **b** a la cual la probabilidad de formación es p=1/2 (hallado mediante un algoritmo), un parámetro de homofilia **alpha** que modula la formación, y una medida de distancia **d()**.

1. ¿En qué se relaciona con mi tema de investigación?

Permite generar las redes small-world con aquellas propiedades vistas en redes sociales reales.

*Therefore homophily may play an important role in the processes generating social networks for it reproduces the crucial properties of clustering and assortativity and at the same time has a natural sociological interpretation as well as has been frequently observed in empirical data.*

*…homophily is indeed sufficient to reproduce most of the typical properties of social networks, but it does not imply the small-world property. However, we also showed that even a small amount of random edge rewiring, as done by Watts & Strogatz (1998), is enough to guarantee it without destroying the structure induced by homophily.*

*Moreover, based on our results we proposed a simple rule of thumb for distinguishing between networks that are likely and unlikely to be shaped by homophily.*

**Clustering and positive assortativity 🡪 Homophily can be the leading generating process**

(Assortativity: Degree assortativity quantifies the tendency of nodes to connect to other

nodes of similar degree.)