CESifo WORKING PAPERS

7508
2019

February 2019

Optimization over Graphs

Suehyun Kwon

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7508*
Category 12: Empirical and Theoretical Methods

Optimization over Graphs

Abstract

This paper defines isomorphism for network formation that is not mathematically isomorphic. Once the pattern of network formation is characterized, one can test for (i) location fixed effects, (ii) heterogeneity, (iii) private information allowing for long-term contracts. The paper provides tests for directedness, heterogeneity, isomorphism and different types of stability.

Keywords: networks, location fixed effects, path dependency.

Suehyun Kwon
Department of Economics
University College London
30 Gordon Street
United Kingdom – London, WC1H 0AX
suehyun.kwon@gmail.com

February 5, 2019
First draft: April 11, 2018
I thank Larry Samuelson for comments.

1 Introduction

This paper studies optimization over graphs, and in particular, nonparametric estimation allowing for location fixed effects within a network. Depending on what types of networks one is studying, one might want to allow for long-term contracts instead of habit formation. For example, in any supply chain, some companies change their suppliers quite frequently, but some don't. Consider the computer industry and the suppliers of CPU, GPU, RAM, graphic board and so forth. In particular, when it comes to semiconductors, there are very few firms producing them in the first place; or to be more precise, there are very few who can design a chip. When a major hardware company changes the CPU supplier, it is announced months in advance that starting this model that will be released from this date, the CPUs will be supplied by a different firm.

I first characterize the pattern of network formation literally only taking into account nodes and links in each date. Think of each node as a firm and each link as a contract. In addition to isomorphism in the sense of graph theory, one can identify non-isomorphic networks that exhibit the same pattern of network formation. The latter is what I call "economic isomorphism"; as graphs, they are not the same, but their behavior of forming links is the same. Once these patterns are identified, we can start asking why some networks are only economically isomorphic and not mathematically isomorphic.

There could be many reasons for the isomorphic behavior, and one can test the decision-theory side or the expected utility as well, but I first focus on the location fixed effects. Assuming the standard expected-utility maximizer, if one were to consider homogeneity, heterogeneous characteristics and private information, then location within a graph is a unique feature for networks or graphs. If the shape of the graph is known to all players, then location is observable. And for most nodes within the same graph, their locations within the graph is not isomorphic, i.e., not mathematically equivalent. Then there could be other heterogeneous characteristics, and some might be private information. Types of private information at each node can include any long-term contracts, clauses, exceptions, shocks over time and so forth.

In order to test beyond isomorphism, one can make modelling assumptions and test parameters. Or one can start by ruling out assumptions that don't explain the data. In order to extrapolate or compute the counterfactual, one needs some degree of continuity in the model. When we're testing a parameter, then the objective function needs to be continuous, but when we're testing strategies or private information, one needs to consider when a strategy is continuous in heterogeneous characteristics, private information, etc and also when belief is continuous in information.

But even before that, one can first ask whether network formation depends on any private information at each node and how much the location fixed effect can explain amongst the observable characteristics. These still don't require any particular modelling assumption. But one needs to discuss projection in the vector space and why orthogonality, not independence or covariance zero, is required. For anyone not familiar with the vocabulary, consider the x-coordinate and the y-coordinate of a two-dimensional vector (x, y). And think about drawing the two axis that are not orthogonal to each other. If one were to find the x-coordinate and the y-coordinate as before, one ends up getting two vectors that are not orthogonal to each other. So far, so good. Now at this point, consider $ax+by$ where a, b are the estimates and x, y are the independent variables. If x, y are not from orthogonal axis as in the example above, then the estimates a, b depend on the order of projection in the estimator; consider the following two cases.

$$\begin{aligned}(1,1)+(\frac{3}{2},0) &= \big((1,0)+(0,1)\big)+\frac{3}{2}(1,0)\\ &= (1,1)+\frac{3}{4}\big((1,1)+(1,-1)\big)\end{aligned}$$

On the other hand, when there are too many independent variables, one can't identify anything, but in a graph, if one were to consider the possibility of location fixed effects, the number of "location" is already large. I provide a test whether location fixed effects can explain the data without further heterogeneous characteristics, private or public. At this point, one needs to discuss why one should consider a network instead of treating the dataset as usual. If each contract is exclusively bilateral, then the problem can be considered as matching instead of networks. However, most companies have multiple suppliers, and some sign exclusive contracts, but some don't. Does it lead to any loss of generality if one shuts down any location fixed effect without testing?

In operations research and mathematics, (specifically combinatorics) graphs have been studied for decades, but it hasn't been as long in economics. In economics, most of the times, networks refer to graphs, but some papers only take the average number of links, and other papers also make specific assumptions about the model. Most recently, there is de Paula, Richards-Shubik, Tamer (2018) that is published in Econometrica.

Within economic theory, network formation hasn't been studied extensively, and network formation with perfect foresight was pretty much non-existent until four years ago. Stability is a concept that applies to an existing graph or a network, and most papers in the literature have focused on short-sighted agents when it comes to network formation. To put it differently, given the set of nodes and links, if an agent at a particular node is going to form a new link, who is he going to form the link with? It depends on several factors, but one of the key factors is whether the agent has perfect foresight, i.e., he knows the final stable network that will be formed allowing for mixed strategies. Other factors include whether the agent knows the entire network structure, the exit probability, the state of the world, the cost of link formation among others.

My algorithm classifies all graphs within the data set.[1] It allows for both directed graphs and undirected graphs, but the key is to classify the pattern of network formation in the data set. If the data set is dynamic, one can test whether each graph is stable, but a weaker condition is whether the distribution of graphs is stable. With a dynamic data set, one can test it directly, and an even weaker condition is whether the distribution of different "classes" of graphs is stable.[2] But the last condition depends on the classes of graphs

[1]de Paula et al only extracts pairwise stable networks.

[2]Classes refer to a partition of the set of all graphs for the given (finite) number of nodes.

which can depend on each application. If the data set is only static, one can analyze it under the assumption that the distribution is stable. Having a panel data can help identifying appropriate classes of graphs. None of my results require any stability or stationarity a priori, and classification can be considered as justifying the data set as a steady distribution under the given network formation and classes of graphs.

At an abstract level, my approach treates each graph or network in the sample as a data point in the "mega" network. Each graph type corresponds to a node, and any graph that can be reached by adding a node or a link is connected by a link from the initial graph. In principle, this network over graphs is completely determined by the number of nodes, but the data set provides the probability weight of each node, and one can estimate the probability weight of each link between different graph types. The final goal is to classify graphs based on the behavior of network formation in the data. One can compare the network over graphs to the second-order beliefs over the first-order beliefs.

de Paula et al is parametric, and they estimate parameters for preferences with quadratic optimization. They only study pairwise stable networks. My approach doesn't make any assumption on whether heterogeneous characteristics or location within each network leads to different behavior. I can discuss the specific optimization method later, but given a data set, one can ask how to distinguish the heterogeneous characteristics of an agent at a particular node from the location fixed effect within a network. Particularly, if one were to assume that the location fixed effect can depend on each application and that path dependency which is often the case in the theory literature can be present in the data set. This is why I don't restrict attention to stable graphs. When a network is forming, by definition, the graph is changing over time until it reaches a stable one.

The question of how to distinguish heterogeneous characteristics from location fixed effects becomes more problematic with network formation. The location can be endogenously determined, and the heterogeneous characteristics could be part of the reason a particular agent is located at the given node.

In this regard, there can be a causal effect from characteristics to location, but several agents with the same characteristics can be located at different nodes, and what happens afterwards is location fixed effects. If one were to compare the second point to randomized control trials, one can test for location fixed effects with enough number of data points. The optimization method, the limit result and the confidence interval are in three paragraphs. Given that I'm estimating a partition, the definition of point identification needs to be modified.

Whether to assume the distribution of graphs is stable also needs more discussion. As mentioned already, one cannot focus on stable graphs for network formation, but also at the same time, location fixed effects exist independently of whether the distribution is stable. The location of a node within the graph keeps changing, and the location fixed effect can be transient in which case it won't be captured in the steady state. I don't assume any stability a priori, and when the distribution can be supported as a steady state with given network formation and classes of graphs, it is under the assumption that the aggregate distribution is stable. One can further ask whether there should be any nontrivial classes of graphs apart from putting every graph in the same class, what assumptions can be tested, and whether there is a unique partition justifying the distribution. The benefit of identifying partitions that justify the distribution is that one can actually estimate the location fixed effects. Otherwise, if each node in a particular graph has a utility, one can drop the assumption of stable distribution. However, if one is willing to assume the distribution over the partition is stable, this approach doesn't require estimating any utility or ordinal measure in the first place, and this is why one partition can be used to test several models including perfect foresight, myopia and habit formation. One can also think of it as some type of robustness that only requires that the entire network or the whole system is stable.

Compared to the diffusion centrality in Banerjee, Chandrasekhar, Duflo and Jackson (2013), my notion of location fixed effect has a few differences. Their definition relies on the extent information spreads over the network from a particular node which they refer to as information is planted at a node. My definition relies more on isomorphism and concepts in graph theory. The partition itself is estimated from data, but the estimation is over the shape of graphs. The test for heterogeneity is with respect to permutation over nodes, and the notion of stable distribution over the partition of set of all graphs is also new. By definition, location fixed effect assumes that the location with respect to the entire network matters, and the diffusion centrality focuses on the number of steps that the information propagates. This also requires that whether each agent knows the entire network structure, which has become more popular in network theory, also matters. In the network literature, star shapes, completely connected graphs and a few popular shapes have been studied commonly, but there isn't a systematic way of classifying graphs in a few categories if the number of nodes gets large.

Path dependency in the earlier discussion comes from fully rational agents with perfect foresight whose locations are subject to noises (or mixed strategies). Given the empirical distribution and partitions of the set of graphs, one can easily test for perfect foresight and myopia. On the other hand, there are papers in behavioral economics studying habit formation. Given the network of graphs, one can test for habit formation, and different versions of habit formation can be tested by modifying the algorithm for each version. This will move away from nonparametric estimation but can help comparing the fit of each model.

In terms of optimization method, de Paula et al is quadratic optimization. Locally, quadratic optimization can work well, but to estimate a partition of the set of all graphs for the given number of nodes, quadratic optimization might not be the best method. To be more precise, when the probability of each link can be any number between 0 and 1, estimating the partition needs to allow for mixed strategy, and we need a multivariate normal distribution in the limit.

This paragraph is a digression. There are other fields studying estimation over graphs or networks, but there are a few differences that make it difficult to apply their results directly. In computational biology or neuroscience, one can imagine cells are already located beyond certain age, and there cannot be a direct link from the prefrontal cortex to hypothalmus by a synapse. Location choice by an agent is not present in this field. In physics, one can imagine that the existence of node is only probablistic in any instant, but I treat all nodes as either in or out of the network, and the exit probability applies at the end of each period in the dynamic model. In traffic control, the type of network is a directed graph, but it has a particular structure, and methods developed in other fields don't apply to my model immediately. Most importantly, location choice, location fixed effect and perfect foresight don't apply to applications if each node doesn't correspond to agents, and it is even more difficult to test habit formation or any behavioral traits in network formation.

The rest of the paper is organized as follows. Section 2 presents the model, and results are in Section 3. Section 4 concludes.

2 Model

This section describes the model. Given a static dataset, each agent corresponds to a node, and any interaction is a link. The list of all nodes is a vector $\vec{A}$, and the entire dataset is a matrix M. I allow for directed graphs, and whether a graph is directed or undirected can be tested by symmetry of the matrix; entry $M(i,j) = 1$ denotes that there is a directed link from node i to j. Otherwise, it's 0.

If the dataset is dynamic, vector $\vec{A}$ still collects all agents through all time periods in the dataset, but the matrix has a time dimension; $M(t;i,j) = 1$ denotes there is a directed link from i to j in period t. I assume the time is discrete.

When a link can be disconnected, it only changes the entry in M. With entry and exit of agents, a node might appear or disappear at the beginning of period, and one needs to distinguish a node that is not connected to any other node and a node that denotes a node that has disappeared or hasn't entered yet; in a given period, any such non-existent node and corresponding links are denoted by $\cdot$.

Once the dataset is represented by a matrix, the rest is estimation as described in the next section. The source code in Python is in the appendix together with the description of the dataset I tested on.

3 Results

I will first explain basic graph theory in section 3.1 then present main results in section 3.2. Anyone familiar with graph theory can skip section 3.1.

3.1 A Primer on Graph Theory

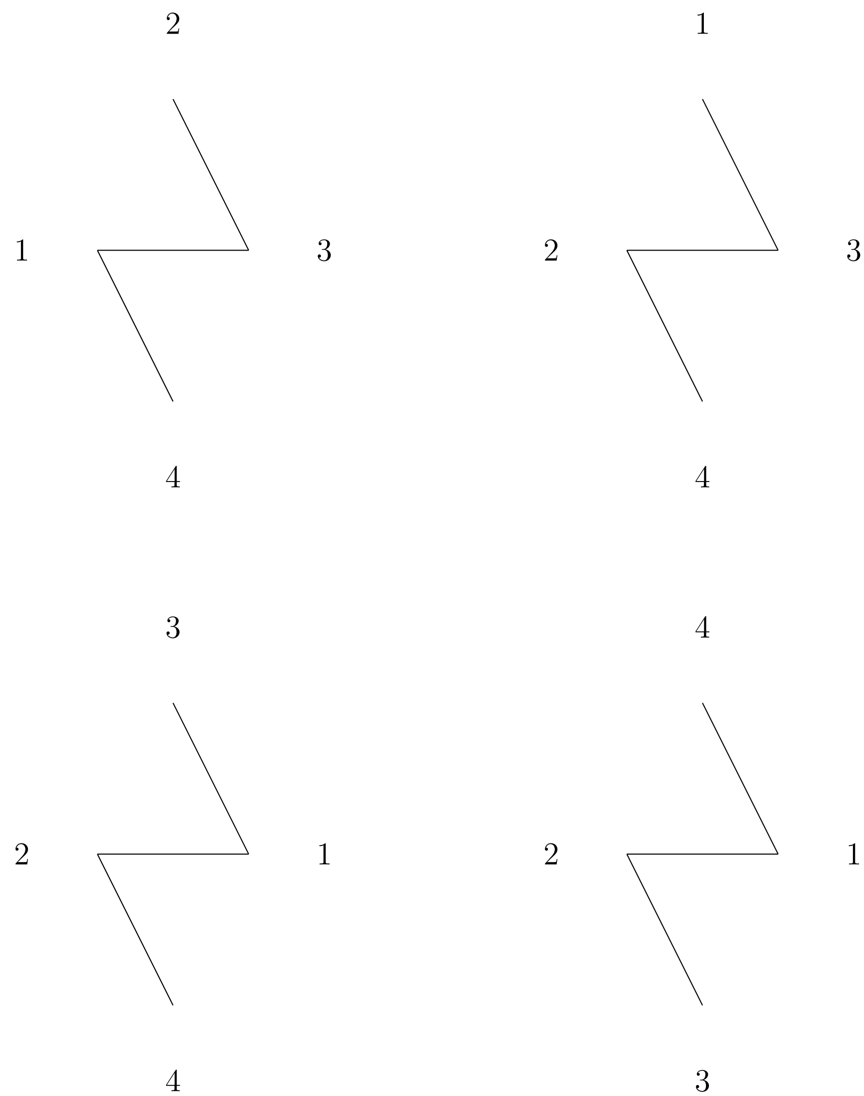
Definitions in this section are permutations and isomorphisms. The concepts do come up in graph theory, but anyone familiar with topology or number theory should also be familiar with these concepts. The most important concept is isomorphism, and given the finiteness of the model, ring isomorphism is the underpinning concept in this paper. I apply isomorphism on individual graphs but also more stringently in section 3.2.

Formally, permutation over a finite set is defined as follows:

Definition 1. *A permutation over $I = \{1, \cdots, n\}$ is a one-to-one function $f : I \to I$.*

Just to remind what one-to-one function means, this means that for each $i \in I$, there is only one $j \in I$ such that $f(i) = j$. Likewise, for each $j \in I$, there is only one $i \in I$ such that $f(i) = j$. There are altogether $n! = n \times (n-1) \times \cdots \times 1$ permutations over the set of I when $|I| = n$. To prove it, there are n elements in I that can be $f(1)$, then by the definition of one-to-one mapping, there are $n-1$ elements that can be $f(2)$. This leads to $n+1-k$ possibilities for $f(k)$. Since each of these choices are independent, one needs to multiply all of them.

I first illustrate isomorphism with graphs then define more formally. Consider the first graph with 4 nodes. There are $24 = 4 \times 3 \times 2 \times 1$ permutations over $(1, 2, 3, 4)$, but if you consider $(2, 1, 3, 4)$, this only swaps nodes 1 and 2: this is the second graph. Another permutation is cycling through $(1, 2, 3)$, and if you consider $(2, 3, 1, 4)$, the third one is the corresponding graph. Finally, when all nodes are permuted, an example is $(2, 4, 1, 3)$:



All four graphs "look" the same, and the only difference is the "label" at each node. From definition 1, when $|\vec{A}| = n$, there are $n!$ permutations altogether. The number of graphs depends on the types of graphs we're considering. One can allow for directed graphs or undirected graphs. Then one can also consider a link from a node to itself or not allow such links. In case anyone is familiar with the notation $\binom{n}{k}$ or ${}_nC_k$, it is straightforward to count the number of graphs. Otherwise, read through the rest of this paragraph. The number of graphs is 2^{n^2} or $2^{n(n-1)}$, since between each pair of nodes, there are two possibilities; there is a link, or there isn't. The first case, 2^{n^2} allows for a link from a node to itself, but it can be dropped. Out of $2^{n(n-1)}$ different graphs, allowing for $n!$ permutations, one can characterize the number of different "shapes" of graphs. When two graphs "look" the same, they are isomorphic to each other, and with the type of graphs in this paper, i.e., finite and non-stochastic, the only isomorphism between individual graphs is permutation over nodes. Or more precisely, given a graph, any isomorphic graph relabels all the nodes while keeping the diagram the same.

The reason I'm counting the number of graphs is to illustrate how many different types of graphs we get if we treat all isomorphic graphs as one type. In the example above, I drew four different graphs that look the same. So are these isomorphic or not?

I will leave the formal definition of isomorphism for graphs with a finite number of nodes and links as graphs that look the same except for the name or the label of each node and link. Note that there is a qualifier "finite number of nodes and links." The word "isomorphism" can mean different things even within mathematics, depending on the object you're discussing, and I won't get into discussions of non-relevant definitions here. Later in the paper when we discuss probabilities of forming a link, these can be any number between 0 and 1, but in any given period, either there is a link or there is no link between any two nodes. Therefore, graphs themselves in the dataset have a finite number of nodes and links.

The four graphs I drew earlier look the same, and they are isomorphic to each other. I also said that there are altogether 24 permutations of $\{1, 2, 3, 4\}$. Given that the examples are undirected graphs without any link from a node to itself, let's count the number of isomorphic classes of this type of graphs with four nodes. Undirected graph means that we just need to count the number of pairs of nodes, because each link has to connect two nodes; with four nodes, we have $(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)$. For each one of 6 possible links, there may or may not be a link, and therefore, the total number of undirected graphs is $2^6 = 64$.

Next, how many isomorphic classes are there? Before counting the number of isomorphic classes, I'll state the following fact. It states that only isomorphic graphs with a finite number of nodes and links are the ones whose nodes are permuted.

Fact 1. *The only isomorphism over individual graphs when each graph has a finite number of nodes and is permutation over nodes.*

We know there are 24 permutations and 64 undirected graphs with $\{1, 2, 3, 4\}$. How many isomorphic classes do we have? Below or above 24? Below or above 64? When there is 0 link, there is only one graph. When there is 1 link, there are 6 isomorphic ones. When there are 2 links, either there is a common node in both links (12 isomorphic ones) or two links are separate (3 isomorphic ones). Before counting the cases with 3 links, we can flip the graph by (i) if there is a link, the new graph doesn't have a link between the pair of nodes and (ii) if there is no link, then there is a link in the new graph. Therefore, we know there is only one graph with 6 links, 6 isomorphic ones with 5 links, 12 isomorphic ones and 3 isomorphic ones each for 4 links. This leaves 20 graphs with 3 links, and the isomorphic classes are (i) all three links share the same node (4 isomorphic ones), (ii) three links form a chain (12 isomorphic ones), and three links form a triangle (4 isomorphic ones). Note that the way I counted the number of graphs treat two mirror images as one.

3.2 Estimation Methods

This section shows the estimation methods together with intermediate tests. The first test is symmetry, or whether the network is directed or undirected. An undirected graph is equivalent to each link being directed in both directions for each pair of nodes.

Test 1 (Test for directed graphs)**.** *Let* $d(i, j) = M(i, j) - M(j, i)$ *for all* $i \geq j$*. When the data set is dynamic, define* $d(t; i, j) = M(t; i, j) - M(t; j, i)$*. The following three statements are equivalent to each other and imply the fourth.*

1. The matrix M is symmetric.

2. The corresponding graph is undirected.

3. The network flow is 0 over every link.

4. The network flow is 0 at every node.

When M is not symmetric, $d(i,j)$ or $d(t;i,j)$ keeps track of the network flow.

The network flow is related to transportation problems, but transportation over manifolds is different from discrete non-stochastic graphs which belong to combinatorics to be strictly speaking. Another difference is that in Theorem 1, the equivalent statement is for the network flow over each link, not at each node. One can characterize further relationships between two definitions as shown in the fourth point, but Hassin (1982) or network flow in the operations research literature is mostly about network flow at each node. If one were to test whether the distribution of graphs is stable, one can still allow for permutation over nodes. Graphs can cycle through different partitions (classes), and the network flow by itself doesn't have to be 0 at any node or any link when the distribution is stable. It characterizes a different condition, and one doesn't imply the other.

For the rest of the paper, the number of graphs, the number of links, the number of isomorphisms and so forth are for undirected graphs, but everyting else works for directed graphs verbatim.

The following test only applies to dynamic models. It is a preliminary test for stationary distributions without classifying graphs. As mentioned in the introduction, classifying graphs is the same as identifying the partition of the set of all graphs give $|\vec{A}|$. The rest of the paper focuses on dynamic data sets and uses permutations over nodes, which is the only isomorphism over the set of graphs in this paper, i.e., there are a finite number of nodes, and in each period, either there is a node, or there isn't, between any pair of nodes.

Theorem 2 tests between any two time periods, if there is a permutation over nodes such that $d(t;i,j)$ maps into $d(t';i,j)$ as a one to one mapping.

This tests for isomorphism of the set of all graphs, where each graph is treated individually. However, compared to definition **??**, which is between two individual graphs, theorem 2 tests for one permutation that maps the set of all graphs to itself where each graph is mapped to an isomorphic one. As mentioned earlier, stability or stationarity can be relaxed in several steps. Instead of pairwise stability or coalition-proofness, which by definition doesn't apply during network formation, I test for the distribution over all graphs then relax the concept to distribution over partition of the set of all graphs. Theorem 2 tests for the distribution over all graphs, and the distribution over partition is tested in Theorem 4.

Test 2 (Test for stationarity)**.** *One can test the stationarity of distribution up to permutation as follows:*

1. *In each period, order all $\frac{1}{2}n(n+1)$ entries of d in descending order.*

2. *Compare along the time dimension for a permutation of nodes, i.e., for each t, $\{d(t;i,j)|i \geq j\}$ with $\{d(t+1;i,j)|i \geq j\}$.*

3. *If there exists any t such that there is no permutation between the set of all graphs in period t and $t+1$, the distribution is not stable up to permutation, when each graph is treated individually.*

When the number of nodes is not too big, one can compute CDFs of $d(t;i,j)$ for all permutations and test for stationarity. An alternative is to order $d(t;i,j)$ without computing CDFs, but there is $n!$ permutations when $n = |\vec{A}|$. The methodology itself doesn't change with the number of nodes, but the algorithm can be improved for efficiency. The most efficient way is to order all $\frac{1}{2}n(n+1)$ entries of d, then compare along the time dimension for a permutation.

Given the first two tests, the next step is to estimate the partition if the distribution with each graph as an individual partition is not stable. This part depends on the data and each application to be more precise. As mentioned in the introduction, a panel data makes it the easiest, and with a static data set, one should assume that the distribution is stable.

When the number of nodes is finite, the number of graphs, the network over graphs, and all possible partitions of the set of all graphs are all finite. What I refer to as estimating the partition of the set of all graphs is not that one can always just pick one or divide a partition into two among the finitely many existing candidates. The goal is to identify the pattern of network formation, and if any agent uses a mixed strategy, then the data set can no longer be treated as isomorphic between any two periods. Furthermore, the key here is that in addition to two obvious partitions, i.e., putting all graphs in one or putting each graph into a separate partition, in terms of location fixed effects, there can be nontrivial partitions of graphs.

Before characterizing a partition in Theorem 4, I first test for heterogeneous characteristics in Theorem 3. Permutation over nodes, which is the only isomorphism, is another trivial partition, and this partition provides a test whether nodes can be treated as homogeneous or one should introduce heterogeneous characteristics and disentangle them from location fixed effects. Theorem 2 already tested whether the entire distribution over graphs is stable up to permutation. If the test confirms permutation, then between any two periods, there exists a permutation that maps each graph to an isomorphic one, and the distribution of graphs is identical throughout the entire time period in the data set. If one were to consider the network over graphs, then this network is isomorphic between periods t and $t+1$.

Theorem 3 tests whether the pattern of network formation is stable up to permutation. Compared to Theorem 2, Theorem 3 doesn't require that the entire distribution is isomorphic between every consecutive periods.

Test 3 (Test for heterogeneity)**.** *One can test for heterogeneity as follows:*

1. *Group all graphs that are isomorphic up to permutations of nodes in the first period.*

2. *For each case, test whether the subsequent behavior are isomorphic up to permutation, i.e., between the set of graphs that can be reached from one graph and the set of graphs that can be reached from an isomorphic graph,*

there exists a permutation of nodes that is a one-to-one mapping between two sets, and furthermore, each graph is mapped to an isomorphic one.

3. *When mixed strategies are allowed, one can provide the error size in the probability of each link for the given sample size.*

If the hypothess is rejected, there is no heterogeneous characteristics, and the only difference comes from mixed strategies.

I haven't provided the confidence interval in Theorem 3, but the limit result as the sample size grows to infinity can be shown easily. However, due to the number of different types of graphs for the given number of nodes, the test requires a higher number of data points than some of the more common estimation methods to confirm the hypothesis.

In the rest, I assume the test for heterogeneous characteristics has failed and treat all nodes as homogeneous. In estimating any nontrivial partition, one needs to define the "location" of node when the graph itself is not stable. One can always compare isomorphic ones, but when two graphs are not isomorphic, this requires an additional definition. Most importantly, the graph is changing, and one needs to define the location fixed effect with respect to the shape of the graph in each period. The last point also requires an assumption as how to treat the location fixed effect over time. An obvious approach is to make it additively separable over time as with usual discounting, and any persistent effect is captured through the new location in the following period. This approach needs to take care of the problem that the location of a node within the network changes over time, but the effects are not necessarily independent. However, one can consider this as orthogonalizing the effect over time, and once a new location is determined, the new location picks up on all residual effects. This can be microfounded if necessary, but it either requires an assumption on behavior or strategy; Markovian strategies can be optimal with perfect foresight which provides one microfoundation for the orthogonalization assumption.

Once the partition is identified, then one can test perfect foresight and different models of habit formation, including the degree or number of steps of foresight. Also, the partition itself provides location fixed effects in the data set which is the main contribution of Theorem 4.

Before preceding, one should note that Theorem 4 doesn't assume anything about the agent's optimization problem at each node. This is another reason my approach is different from de Paula et al or any other parametric papers. Theorem 4 just characterizes partitions of the set of all graphs given the data set. However, the remaining theorems test the agent's optimizaton problem, even allowing for non-expected-utility maximizers. The key step for this theorem is partitioning the set of all graphs and defining the location fixed effect.

Test 4. *Suppose the distribution of graphs is not stable and the test for heterogeneous characteristics failed. One can find all partitions of the set of graphs that justify the distribution as stable as follows:*

1. *There are* $2^{\frac{n(n+1)}{2}}$ *undirected graphs or* 2^{n^2} *directed graphs.*
2. *Set aside two trivial partitions: one that includes all graphs and the other that treats each graph individually (already tested in Theorem 2).*
3. *Group all graphs that are isomorphic up to permutation of nodes.*
4. *Identify all nontrivial partitions according to Definition 2 below.*

As mentioned already in the introduction, the diffusion centrality in Banerjee, Chandrasekhar, Duflo and Jackon (2013) measures the number of steps information spreads through an existing network. The network itself doesn't change, and the information propagates through the network. Location fixed effect in my model measures the effect of earlier locations on the final location once the network is formed. It does depend on the number of immediate links, once removed, twice removed and so forth, and it is related to the diffusion centrality, but the effects on information transmission and network formation don't need to be the same. However, there is a common feature between the two: given the data set, the diffusion centrality is estimated immediately from the data, and the location fixed effect is estimated by the shape of graphs in the data. Both depend on the data set and the application, and obviously, the existing network and the number of links matter for both. Once they are estimated, one can microfound the correlation between the two which should come from the agents' utility function together with the degree of rationality or foresight.

Strictly speaking, isomorphism only applies to permutation of nodes. However, one can define a correspondence from the set of all graphs to itself given the number of nodes n, and with pure strategies and mixed strategies, the partitions we're looking for are subsets whose image of this correspondence is in one subset.

Definition 2. *Let* $N = 2^{n(n+1)}$ *be the number of undirectd graphs given the number of nodes* n*. Denote by* S *the set of all undirected graphs. A partition* $\mathcal{P}$ *is a collection of nonempty disjoint subsets of* S *whose union is* S*.*

$$P = \{P_1, , P_k\} \ s.t. \ \emptyset \neq P \subseteq S, \ P_i \cap P_j = \emptyset, \cup_i P_i = S.$$

When at most one link can be added in a given period, a partition $\mathcal{P}$ *justifies the empirical distribution* F *if for each* i*, there exists* j *such that*

$$f(P_i) \subseteq P_j$$

every period, and $f(s)$ *is the set of all graphs that can be reached by adding a link from* s*.*

When every node in a graph can form links using mixed strategies in any given period, a partition $\mathcal{P}$ *justifies the empirical distribution* F *if there exists a correspondence* f *such that* $f(s)$ *is a nonempty subset of all graphs that can be reached by mixed strategies at each node in a given period, and for each* i*, there exists* j *such that*

$$f(P_i) \subseteq P_j.$$

First of all, correspondence is a mapping where $f(s)$ for each s is a subset instead of a single element as with functions.

The number of partitions is $1+2^{N-1}+\cdots+1$. Even if strategies are mixed, the graph in each period is not stochastic, and the data provides the empirical mixing probabilities. I focus on usual equilibrium notions without correlated signals, and this implies that the probability of each link is independent from each other; we'll get a multivariate normal distribution in the limit. This is different from Chandrasekhar-Jackson (2017), but the definition can be generalized in a straightforward manner to allow for correlation. The off-diagonal entries won't be 0 in the multivariate normal distribution in the limit. One should also note that it is possible to allow for private information and Bayes correlated equilibrium in principle. This theorem only deals with partitions that justify the empirical distribution so there's no need to pin down the exact equilibrium strategies at this step. But the earlier discussion of orthogonalizing the location fixed effect still needs to be taken into account, and with perfect foresight, optimality of Markovian strategies provides one microfoundation.

With mixed strategies, one can always choose a coarser partition, but finer partitions provide more information. If the correspondence f puts all graphs that can be reached with a positive probability in its image, this characterizes coarser partitions, and one can define finer partitions. And this is precisely why there is no reason to think that with mixed strategies, all partitions justifying the same empirical distribution can be nested or there exists the finest partition nested by all other partitions. Empirically, when one partition is a refinement of the other, obviously any model that explains the behavior of a finer partition also explains the coarser partition, but unless the sample size is sufficiently big, it is unclear why a refinement should be the best explanation. The definition requires that instead of collecting all graphs that can be reached in a given period, a partition is defined jointly with a correspondence.

The main difference of estimating a partition that makes the aggregate distribution stable in a dynamic data set from other approaches is that this approach doesn't require any ordinal measure. The data set consists of the pattern of network formation, but most empirical papers on networks also have data on nodes and links, but they don't have data on each agent's utility function or all relevant information directly. Otherwise, they need to be collected which requires more work on the data collection method; particularly, during network formation, one can always miss a node that enters later, and the existence of a link can be obtained far more easily a few periods later than any utility level. My method agrees with the existing method in a sense that each link is coded as a binary variable. But I focus on the shape of graphs during network formation instead of other behavior on already stable networks. Compared to parametric estimation, my estimation only requires the assumption that the aggregate distribution is stable with respect to some partition. This is an assumption, but compared to steady state assumption in macro, it's not a very stringent assumption. Furthermore, this is nonparametric in a sense that it doesn't use any assumption on the agent at each node. The only thing that is needed is whether there is a link or not between any two nodes in a given period. Therefore, if one were to assume the aggregate distribution is stable, the estimated partitions can be used to test any candidate behavior of agents.

Once the potential partitions of the set of all graphs that justify the aggregate distribution as stable are identified, the rest is comparing different distributions over the set of all graphs and is relatively straightforward. It takes more work to compute the theoretical distribution that needs to be compared to the empirical one.

4 Conclusion

I study the pattern of network formation in this paper. I characterize a method to analyze the data set in three main steps. After testing whether the graphs are directed, I test if the aggregate distribution of graphs is stable in the panel data. I provide a test for heterogeneous characteristics, but it requires a relatively large sample size to confirm the hypothesis, and the rest of the paper focuses on the location fixed effect while treating all nodes as homogeneous.

The key step in test 5 is the definition of location in a given network and the location fixed effect when the graph is not stable. If one were to treat nodes as heterogeneous, one needs to disentangle the heterogeneous characteristics and location fixed effects while the network is formed. I focus on location fixed effects with homogeneous nodes, but this still allows for mixed strategies. Probably the closest or the most relevant concept to location fixed effect is the diffusion centrality in the microfinancing randomized control trial literature.

I estimate partitions of the set of all graphs, and one could test for myopia, perfect foresight, and different versions of habit formation in the literature. The first two cases still allow for path dependency, and this can be applied to randomized control trial data sets. In principle, one can run these tests without estimating partitions. However, these partitions depend on each data set and each application; when the distribution of partitions is stable, each partition reflects economics of a particular data set or an application. Since the estimation doesn't require any assumption about agents, partitions can be used to identify the objective function of agents. When all nodes are treated as homogeneous, all differences are attributed to location fixed effects, and as is always the case with networks, if one were to put an independent variable to each data point, one can't estimate anything. Nontrivial partitions show that this doesn't have to be the case, and they provide insights into how to define locations for a particular application. It automatically follows that the approach doesn't work if one were to put one heterogeneous characteristic for each data point. With parametric estimation, there are specific assumptions about the utility function, but the support for each variable is continuum. One needs to take caution as location fixed effects might be attributed to heterogeneous characteristics, particularly the effect of initial location.

I don't disentangle heterogeneous charactetistics from location fixed effects in this paper, but to test these two properly, one might have to focus on new nodes and new links, and randomized control trials might be better than other types of data sets. Once a network is already formed, any existing node or link already reflect both.

References

[1] Chandrasekhar, Arun, and Matt. O. Jackson. (2017) "A Network Formation Model Based on Subgraphs," working paper.

[2] Banerjee, Abhijit, Arun Chandrasekhar, Esther Duflo, and Matt. O. Jackson. (2013) "The Diffusion of Microfinance," *Science,* 341(6144) DOI: 10.1126/science.1236498.

[3] de Paula, Aureo, Seth Richards-Shubik, and Elie Tamer. (2018) "Identifying Preferences in Networks With Bounded Degree," *Econometrica,* 86(1): 263-288.