Importance of Professional Networks in Trade: Evidence from Real Estate Market

Jia Xie, California State University, Fullerton
Importance of Professional Networks in Trade: Evidence from Real Estate Market

Jia Xie
Department of Finance
California State University Fullerton
jxie@fullerton
(657) 278-4059
July 17, 2019

Abstract

A professional network is formed among real estate agents when they trade with each other on behalf of clients. In the process of finding trading partners, agents either search randomly or navigate through the local network. This paper estimates the importance of network-based searches relative to random searches by fitting Jackson-Rogers’ (2007) network formation model to the Multiple Listing Service (MLS) data for a major Midwestern city in the U.S. during 2008-2010. We find that 35%∼55% of trades are made through network-based searches. Moreover, agents rely mainly on random searches at the early stage of career, but more on network-based searches after having many trades. On average, a real estate agent trades with one out of every 5 to 10 agents she finds through random or network-based searches.

Keywords: Real estate agent, network, random search, network-based search.
Importance of Professional Networks in Trade: Evidence from Real Estate Market

July 16, 2019

Abstract

A professional network is formed among real estate agents when they trade with each other on behalf of clients. In the process of finding trading partners, agents either search randomly or navigate through the local network. This paper estimates the importance of network-based searches relative to random searches by fitting Jackson-Rogers’ (2007) network formation model to the Multiple Listing Service (MLS) data for a major Midwestern city in the U.S. during 2008-2010. We find that 35%–55% of trades are made through network-based searches. Moreover, agents rely mainly on random searches at the early stage of career, but more on network-based searches after having many trades. On average, a real estate agent trades with one out of every 5 to 10 agents she finds through random or network-based searches.

Keywords: Real estate agent, network, random search, network-based search.
1 Introduction

A professional network is the collection of mutually-beneficial connections among institutions and individuals. Professional networks are important for business and at workplace; People rely on professional networks to achieve various outcomes such as sharing information on job opportunities, making additional sales, access to talent for recruitment, collaboration among researchers in academia, joint ventures among firms, political alliance among countries, et al.

Professional networks are built over time through participants searching and matching with each other. An interesting characteristic of the search-and-match process is that a participant may find other participants to connect with by randomly meeting people or by navigating the local network. For instance, people may find new friends by meeting people at random or through existing friends, i.e., meeting friends of friends. A researcher may find new coauthors at random, or through their existing coauthors and see who else they have collaborated with.

In this paper we focus on the professional network of real estate agents formed through trading. In the real estate market, the seller and the buyer often do not trade with each other, but instead hire real estate agents to represent them. The agent representing the seller is the listing agent, while the agent representing the buyer is the buyer agent. When a trade is made, a connection or link is formed between the two agents. To find a potential trading partner, a real estate agent often adopt two search strategies: (i) a random search where she meets other agents uniformly at random, and (ii) a network-based search where she meet other agents through her local network. There has been some anecdotal evidence from the industry that the network-based search is important, and some even argue that agents rely more on the network-based search than the random search in making transactions. There, however, has been no research that either identifies or quantifies the relative importance of the network-base search. And this paper tends to fill in this blank.

For the purpose of this paper, we adopt the dynamic network formation model in Jackson
and Rogers (2007) and fit it to the Multiple Listing Service (MLS) data from a major Midwestern city in the U.S. during 2008-2010. The MLS data represent a trading network of 1,588 agents with 3,331 links. We find that 35%~55% of trades are made through network-based searches. Moreover, agents rely mainly on random searches at the early stage of career, but more on network-based searches after they have had many trades. On average, a real estate agent trades with one out of every 5 to 10 agents she meet through either random or network-based searches.

There is a vast literature studying the search behavior of real estate agents. Some studies focus on the impact of MLS on agents’ search effort (Yinger, 1981; and Li and Yavas, 2015b), and the results are mixed: While Yinger (1981) finds that MLS reduces agents’ commissions and search effort, Li and Yavas (2015b) find that introducing an MLS reduces the number of agents but increases each agent’s effort level. Some other studies find that the agency problem—where real estate agents receive only a small fraction of the sale price as commission—lead to second-best effort level (Zorn and Larsen 1986; Anglin and Arnott 1991; Geltner, Kluger and Miller 1991; Fisher and Yavas 2010). Miceli (1991) and Miceli, Pancak and Sirmans (2007) find that the commission split structure—where the listing agent receives a significant portion of the total commission even when a buyer is located by the buyer agent—also reduces agents’ search effort. Yavas (1992, 1994) find that the use of real estate agents reduces the equilibrium search intensities of buyers and sellers. Li and Yavas (2015a) finds that agents’ search effort increases with market tightness (ratio of buyers to sellers). Different from the literature which focuses on the optimal/equilibrium level of the agent’s effort in various circumstances, we instead study the allocation of agents’ effort between random and network-based searches.

This paper also belongs to the literature on social networks. Jackson and Rogers (2007) and Jackson (2008) propose a hybrid network formation model, which combines the random network and the preferential network (Newman, 2003) and is analytically trackable. Later on, Chaney (2014) applies this model to the international trading partner network of French
firms, and finds that network-based search is twice as important as random search for firms with a single foreign contact, and become even more important as firms gain more contracts. In this paper, we apply this model to the trading network of real estate agents and find that network-based searches are also important in the real estate market.

The rest of the paper is organized as follows. Section 2 explains the motivation for the study, Section 3 introduces the dynamic network formation model, and Section 4 introduces the MLS data. In Section 5, we fit the model to the data to get the main results of this paper. Section 6 does some robustness analyses and Section 7 concludes.

2 Motivation

This study is motivated by two observations from the MLS data that certain characteristics of the agent trading network are very different from those of a simulated random network, suggesting that the agent network can not be formed uniformly at random.

To facilitate the comparison, let’s first formally describe the agent trading network as follows. Given a set of real estate agents, $N$, where each agent is denoted by a node $i$, a trading network is a $N \times N$ matrix, $g$, where $g_{ij} = 1$ if there is a directed link from $i$ to $j$, and $g_{ij} = 0$ otherwise. A directed link from $i$ to $j$ represents trades (i.e., real estate transactions) initiated by $i$ and agreed by $j$. The agent trading network can be simulated by a Poisson random network which has the same number of nodes $N$ as in the agent network, and the same number of links, each of which is formed uniformly at random. Next we will compare the two networks along two dimensions and show the differences.

The first difference that we observe between the actual agent network and the simulated random network is on the degree distribution $F(d)$ defined as follow: First, we define the degree of node $i$, denoted by $D(i) = \sum_{j=1}^{N} (g_{ij} + g_{ji})$, as the number of nodes linked to or from $i$, or equivalently the number of other agents with whom $i$ have trades (either to or from). Then $F(d)$ is the fraction of nodes in the network with degrees no more than $d$. That
Figure 1: Complementary distributions of degree for the actual agent network (dot) and for the simulated random network (square)

Notes: This figure shows the log-log plot of the complementary cdf of the degree distribution for the actual agent network in the data (the dots) and that for a simulated random network (the squares), respectively. The figure suggests that the random network model fits the data poorly.

is, \( F(d) = \sum_{i=1}^{N} I(D(i) \leq d)/N \), where \( I(\cdot) \) is an indicator function. Figure 1 provides a log-log plot of the complementary distribution of degrees; the horizontal axis is \( \log(d) \) and the vertical axis is \( \log(1 - F(d)) \). Two complementary distributions are shown here: The dots are for the actual agent network, and the squares are for the simulated random network. As we can see, the degree distribution for the agent network is very different from that of the random network.

The second dimension along which we find differences between the agent network and the random network is the fraction of transitive triples. A **connected triple of nodes** is a relation where \( i \) has a directed link to \( j \) and \( j \) has a directed link to \( k \) (Figure 2). If there is also a link from \( i \) to \( k \), then the connected triple of nodes is also a **transitive triple**. The intuition behind the concept of transitive triples is that “friends of my friends are my friends.” The fraction of transitive triples, denoted by \( C^{TT}(g) \), is the relative number of transitive triples among all the connected triples in the network. In a general sense, \( C^{TT}(g) \) measures how likely agents \( i \) and \( j \) are connected to each other if both are connected to the same agent.
The connected triple of nodes and the transitive triple

\[ C^{TT}(g) = \frac{\# \text{ of transitive triples}}{\# \text{ of connected triples of nodes}} = \frac{\sum_{i:j \neq i; k \neq i, j} g_{ij}g_{jk}g_{ik}}{\sum_{i:j \neq i; k \neq i, j} g_{ij}g_{jk}}. \] (1)

The actual agent network has 16,672 connected triples of nodes, among which there are 484 transitive triples. Therefore, the fraction of transitive triples for the agent network is 0.029. On the other hand, the simulated random network has only 4,639 connected triples of nodes, among which only 13 are transitive triples, implying that the fraction of transitive triples for the random network is only 0.0028, less than one tenth of that for the agent network. Therefore, the comparison on the transitive triple also suggests that the agent network is not formed uniformly at random, but is in part formed through meeting friends of friends.

3 The Model

In this section, we introduce the dynamic network formation model developed by Jackson and Rogers (2007). Jackson and Rogers’ model is powerful in that it captures all the salient features of the agent network mentioned in the last section, yet simple enough to be analytically tractable.

The network is formed over time with new agents joining the network continuously. Assume time is discrete and denoted by \( t \in \{1, 2, \ldots\} \). At each time \( t \) a new agent joins the network and initiates new links to the existing agents. Links are formed through two types of searches: a random search and a network-based search. First, the new agent will do a
random search among existing agents and meet \( m_r \) of them at random, who then becomes “friends” of the new agent. The new agent then sends each friend a trading proposal, each of which will be accepted independently with a probability \( p \). If a proposal is accepted, a trade is made with the friend, who then becomes a “trading partner” or simply “partner” of the new agent. Therefore, the new agent finds \( p m_r \) partners through random search.

Next, the new agent will do a network-based search to reach out to partners of her friends; these are the agents to who her friends previously initiated trades. The new agent will randomly meet \( m_n \) of them, who then become friends of the new agent. The new agent then sends a trade proposal to each friend found through this network-based search, and each proposal will be accepted independently with a probability \( p \). If the proposal is accepted, a trade is made and the friend becomes a partner. Therefore, the new agent finds \( p m_n \) partners through the network-based search. Altogether, the expected number of links initiated by the new agent, or equivalently the expected number of partners found by the new agent, is \( m = p m_r + p m_n \). To ensure a well-defined network formation process, we also assume that the initial network (at \( t = 0 \)) has at least \( m_r + m_n + 1 \) agents, and each of them has at least \( m_r + m_n \) partners.

Note that we have modeled two levels of relationship among agents: friendship and partnership. While the friendship is built through meeting, the partnership is a deeper relationship built from trading. Correspondingly, two networks exit: the friend network and the trading network. The trading network is a subset of the friend network, in the sense that each link in the trading network corresponds to a link in the friend network, but not vice versa. In other words, a partner of an agent is always a friend, but not vice versa. Since friendship is unobservable in the MLS data, in what follows, we will focus on the trading network. We will develop two formulas to characterize the trading network, and use the two formulas for calibration in the next section.

We define the *in-degree* of an agent \( i \) at time \( t \)—denoted by \( d_i(t) \)—as the number of links going to \( i \), or equivalently the number of agents who have initiated trades to \( i \) as of time \( t \).
Then agent $i$ may get a new ingoing link at time $t + 1$ with the following probability:

$$\frac{pm_r}{t} + \frac{m_r d_i(t)}{t} \frac{pm_n}{m_r(pm_r + pm_n)}.$$

(2)

The first term in (2) is the probability that the new agent is linked to agent $i$ through the random search. Note that at time $t + 1$, there are $t$ existing agents, among whom the new agent will randomly choose $pm_r$ as partners. The second term in (2) is the probability that the new agent is linked to $i$ through the network-based search. The sub-term $\frac{m_r d_i(t)}{t}$ is the probability that some agent—who is linked to $i$—is chosen as a friend by the new agent though the random search; note that $\frac{m_r}{t}$ is the probability of a given agent being chosen as a friend through random search, and $d_i(t)$ is the in-degree of $i$, i.e., the number of agents who are linked to $i$. The other sub-term $\frac{pm_n}{m_r(pm_r + pm_n)}$ is the conditional probability that agent $i$ is chosen as a partner by the new agent, given that some agent—who is linked to $i$—becomes a friend of the new agent.

The analysis of the stochastic dynamic model as described in (2) is complicated. Following the literature, we instead analyze the “mean-field” approximation to the model, by assuming that the in-degree of agent $i$ changes over time continuously (instead of discretely) and deterministically at the mean of the probability (instead of stochastically). In particular, define $m = pm_r + pm_n$ as the expected number of links initiated by a new agent, the following formula is a “mean-field” approximation of (2):

$$\frac{d d_i(t)}{dt} = \frac{pm_r}{t} + \frac{pm_n d_i(t)}{tm}.$$

(3)

Following the standard procedure, the above first-order differential equation can be solved as:

$$d_i(t) = (d_0 + rm) \left( \frac{t}{i} \right)^{\frac{1}{1+\tau}} - rm,$$

(4)

Jackson and Rogers (2007) and Jackson (2008) have shown that the “mean-field” approximation is quite accurate.
where \( d_0 \) is the initial in-degree of agent \( i \) when she joins the network, and \( r = \frac{pm_r}{pm_n} \). The numerator of \( r \) is the number of partners of agent \( i \) found though random research, and the denominator is the number of partners of agent \( i \) found through the network-based search. Therefore, \( r \) measures the importance of random search relative to network-based search for finding trading partners in the real estate market.

This dynamic network formation model generates the following two important predictions—one on the in-degree distribution and the other on the clustering coefficient—based on which we can fit the model to data.

**Prediction 1:** Using the mean-field approximation, the in-degree distribution of the agent network has a cumulative distribution function of

\[
F_t(d) = 1 - \left( \frac{d_0 + rm}{d + rm} \right)^{1+r},
\]

for \( d \geq d_0 \) and each time \( t \).

**Proof.** See the Appendix.

**Prediction 2:** Using the mean-field approximation, the fraction of transitive triples in the agent network is

\[
C_{TT}(g) = \begin{cases} 
p \frac{1}{m(1+r)}, & \text{if } \frac{p}{r} \leq 1, \\
p \frac{m-1}{m(m-1)(1+r)-m(\frac{p}{r}-1)}, & \text{otherwise.}
\end{cases}
\]

**Proof.** See Theorem 2 in Jackson and Rogers (2007).

In Jackson and Roger’s model that we’ve introduced above, the assumption of directional links is restrictive, but it is necessary for the model to be analytically tractable. If we instead assume links are non-directed, then the term \( pm_r + pm_n \) in Equation (2) will be replaced by the degree of the friend of the new agent. The degree of the friend, however, is related to the friend’s age and connections, and therefore can not be treated as a constant anymore. This makes the model much more complicated.
4 The Data

We use the Multiple Listing Service (MLS) data for the suburban areas of a large Midwestern city in the U.S. This data set covers all the residential properties sold by real estate agents during 2008-2010 in that area. For each transaction, the data record the unique ID numbers for the listing agent and the buyer agent. We drop observations with missing IDs of the listing and/or the buyer agents. We also delete dual agent transactions—where the same agent represents both the seller and the buyer—to focus on the connection between different agents. 478 (12%) observations are deleted. The final data contain 1,588 active real estate agents with 3,331 links. On average, each agent has 4.588 trades and 4.19 partners, and therefore 1.094 trades per partner.\(^2\) These agent network characteristics are reported in Column (1) of Table 1.

<table>
<thead>
<tr>
<th></th>
<th>(1) Full Sample</th>
<th>(2) Subsample of experienced agents</th>
<th>(3) Subsample of inexperienced agents</th>
<th>(4) Subsample of “new” agents</th>
</tr>
</thead>
<tbody>
<tr>
<td># of agents</td>
<td>1588</td>
<td>154</td>
<td>1434</td>
<td>179</td>
</tr>
<tr>
<td># of links</td>
<td>3331</td>
<td>787</td>
<td>820</td>
<td>102</td>
</tr>
<tr>
<td>Average # of trades</td>
<td>4.588</td>
<td>24.766</td>
<td>2.421</td>
<td>1.140</td>
</tr>
<tr>
<td>Average # of partners</td>
<td>4.195</td>
<td>21.130</td>
<td>2.339</td>
<td>1.140</td>
</tr>
<tr>
<td># of trades per partner</td>
<td>1.094</td>
<td>1.172</td>
<td>1.035</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note: Experienced agents are defined as those who had more than 10 trades during the data period. The remaining agents are defined as inexperienced agents. About 10% of the agents in the data are experienced. “New” agents are defined as those who had no trades in 2008, the first year of the data period.

5 Fit the Model to the Data

In this section, we calibrate the model with the MLS data. The data do not show who—the listing agent or the buyer agent—initiated each trade, therefore we have to make an assumption on that. Three alternative assumptions can be made: (i) all trades are initiated \(^2\)4.588/4.195=1.094.
by the buyer agent; (ii) all trades are initiated by the listing agent; and (iii) trades are equally
initiated by the buyer and the listing agents. In what follows, we estimate—under the three
assumptions in turn—two parameters of the model, \( r \) and \( p \), by using the two predictions at
the end of the Model section. We then show that the model fits the data well. All empirical
results are reported in Table 2.

5.1 Assumption I: All Trades are Initiated by Buyer Agents

We first estimate \( r \) by using Prediction 1, and then \( p \) by substituting the estimated value of
\( r \) to Prediction 2. Note that \( p \) is not shown in (5) so can not be estimated jointly with \( r \).
All results are reported in Column (1) of Table 2.

To estimate \( r \), note that the in-degree distribution (5) can be rewritten as
\[
\log(1 - F(d)) = (1 + r) \log(d_0 + rm) - (1 + r) \log(d + rm).
\]
We can estimation the average out-degree \( m \) directly from the data and denote it by \( \hat{m} \). By substituting \( \hat{m} \) into the above equation, we get:

\[
\log(1 - F(d)) = (1 + r) \log(d_0 + r\hat{m}) - (1 + r) \log(d + r\hat{m}).
\] (7)

Therefore, we can run a regression of \( \log(1 - F(d)) \) on \( \log(d + r\hat{m}) \) to estimate \(- (1 + r)\). The difficulty is that \( r \) is also in the regressor \( \log(d + r\hat{m}) \). There are two approaches to solve this difficulty. The first is to use the iterative OLS regressions introduced by Jackson and Rogers (2007). In this approach, we first make a guess of \( r \), say \( r_0 \), and regress \( \log(1 - F(d)) \) on \( \log(d + r_0\hat{m}) \) to estimate \( r_1 \). Then we regress \( \log(1 - F(d)) \) on \( \log(d + r_1\hat{m}) \) and get an updated estimate \( r_2 \). We repeat the regression with the updated value of \( r \) until we arrive at a fixed point, where \( r_n = r_{n+1} \).\(^3\) The estimation results are reported in Panel B of Table 2. The second approach is to use the nonlinear least-squares estimation, whose results are reported in Panel C of Table 2. Similar to the iterative OLS estimation, we need to choose an initial value of \( r \). The results turn out to be robust to different initial values of \( r \) and are

\(^3\)We always arrive at the same fixed point of \( r \) with various starting values of \( r_0 \).
similar to those from the iterative OLS regressions.

As shown in Table 2, if all the trades are initiated by buyer agents, then the random-search to network-based-search ratio is 0.746~0.788, implying that real estate agents rely more on network-base searches than random searches; 55.9%~57.3% of trades are made through meeting partners-of-friends.\(^4\) The estimated values of \(p\) imply that the likelihood of trading with a friend during the data period is 10.6\% ~ 10.9\%; in other words, during the data period, an agent will trade with one out of every 9 to 10 agents who she meets during either random or network-based searches.

### Table 2: Estimated Parameters of the Dynamic Network Formation Model

<table>
<thead>
<tr>
<th>Assumptions on who initiated the trade:</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All trades are initiated by buyer agents</td>
<td>2.098</td>
<td>2.098</td>
<td>2.098</td>
</tr>
<tr>
<td>All trades are initiated by listing agents</td>
<td>0.029</td>
<td>0.033</td>
<td>N/A</td>
</tr>
<tr>
<td>Trades are equally initiated by listing &amp; buyer agents</td>
<td>(\frac{1}{1+0.788})=55.9%</td>
<td>(\frac{1}{1+0.746})=57.3%</td>
<td>(\frac{1}{1+0.746})=57.3%</td>
</tr>
</tbody>
</table>

Panel A: Agent Network Characteristics

| \(r\) | 0.788 | 1.890 | 1.569 |
| \(p\) | 0.109 | 0.198 | N/A |
| \(R^2\) | 0.974 | 0.914 | 0.988 |

Panel B: Iterative OLS Regression Results

| \(r\) | 0.746 | 1.694 | 1.201 |
| \(p\) | 0.106 | 0.185 | N/A |
| \(S\) | 0.003 | 0.006 | 0.004 |

Panel C: Nonlinear Least Square Regression Results

Note: \(m\) is the average out-degree of the agent. \(C^{TT}\) is the fraction of transitive triples as defined in (6). \(r\) is the ratio of random search to network-based search. \(p\) is the fraction of an agent’s friends who are also partners of the agent. \(R^2\) is a measure of the goodness-of-fit of the OLS estimation. \(S\) is the standard error of the nonlinear least square regression, which is a measure of the goodness-of-fit of the nonlinear least square estimation.

Next, We evaluate goodness-of-fit of the model. Different measures are used for different estimation methods: For the iterative OLS regressions, we use the \(R^2\) from the last regression to evaluate goodness-of-fit. For the nonlinear least squared regression, since the \(R^2\) is not available, we instead use the standard deviation of the error term, denoted by \(S\). \(S\) is also

\(^4\)1/(1+0.788)=55.9\% and 1/(1+0.746)=57.3\%. 

12
Figure 3: Complementary distribution of in-degree, assuming all trades are initiated by buyer agents

Notes: Both figures show the log-log plot of the complementary cdf of the in-degree distribution for different networks. Figure 3a shows the plot for the data (the dot) and that for the fitted network formation model (the square), respectively. Figure 3b shows the plot for the data (the dot) and that for the fitted random model (the square), respectively. The figure suggests that the network formation model fits the data much better than the random model.

called the standard error of the regression; it measures the average distance that the observed values fall from the regression line. Conveniently, it tells us how wrong the regression model is on average using the units of the response variable, which is the degree distribution $F(d)$. Smaller values are better because it indicates that the observations are closer to the fitted line. Both $R^2$ (for the iterative OLS regressions) and $S$ (for the nonlinear least squared regression) are reported in Column 1 of Table 2. Both values suggest that the model fits the data pretty well. For instance, the $R^2$ from the the iterative-OLS regressions suggest that 97.4% of the actual variation in the frequency distribution of degree can be explained by the model.

In addition, Figure 3a shows the the log-log plot of the in-degree distribution fitted by the network formation model, in comparison with that from the data. It is clear that the model fits the data well, and the fit is much improved from that of the random network, as shown in Figure 3b.
5.2 Assumption II: All Trades are Initiated by Listing Agents

In this subsection, we assume that all real estate trades are initiated by the listing agent. The estimation results are reported in Column (2) of Table 2. The estimated values of $r$ suggest that the random-search to network-based-search ratio is $1.694 \sim 1.890$, implying that $34.6\% \sim 37.1\%$ of trading partners are found through network-based searches.\(^5\) The estimated value of $p$ implies that on average an agent trades with one out of every 5 to 6 agents who she meets through searching.

Also, we can check the goodness-of-fit of the model. First, both the $R^2$ (for the iterative OLS regressions) and the $S$ (for the nonlinear least square regression) reported in Column 2 of Table 2 suggest that the model fits the data well, whichever estimation method we choose. Second, Figure 4a shows the log-log plot of the in-degree distribution fitted by the network formation model, in comparison with that from the data. It is clear that the model fits the

\[^5\]1/(1+1.890)=34.6\% and 1/(1+1.694)=37.1\%.
Figure 5: Complementary distribution of in-degree, assuming trades are equally initiated by listing and buyer agents

Notes: Both figures show the log-log plot of the complementary cdf of the in-degree distribution for different networks. Figure 5a shows the plot for the data (the dot) and that for the fitted network formation model (the square), respectively. Figure 5b shows the plot for the data (the dot) and that for the fitted random model (the square), respectively. The figure suggests that the network formation model fits the data much better than the random model.

Finally, we assume that real estate trades are equally initiated by listing agents and buyer agents. The estimation results are reported in Column (3) of Table 2. The estimated values of $r$ suggest that the random-search to network-based-search ratio is $1.201 \sim 1.569$, implying that real estate agents find $38.9\% \sim 45.4\%$ of their trading partners through network-based search.\footnote{1/(1+1.569)=38.9\% and 1/(1+1.201)=45.4\%}. Since we still do not observe who initiated each trade (at the trade-level), $C^{TT}$ cannot be calculated, neither is $p$. 

5.3 Assumption III: Trades are Initiated Equally by Buyer and Listing Agents

Finally, we assume that real estate trades are equally initiated by listing agents and buyer agents. The estimation results are reported in Column (3) of Table 2. The estimated values of $r$ suggest that the random-search to network-based-search ratio is $1.201 \sim 1.569$, implying that real estate agents find $38.9\% \sim 45.4\%$ of their trading partners through network-based search.\footnote{1/(1+1.569)=38.9\% and 1/(1+1.201)=45.4\%}. Since we still do not observe who initiated each trade (at the trade-level), $C^{TT}$ cannot be calculated, neither is $p$. 

61/(1+1.569)=38.9\% and 1/(1+1.201)=45.4\%.
Also, we can check the goodness-of-fit of the model under this assumption. First, both the $R^2$ (for the iterative OLS regressions) and the $S$ (for the nonlinear least square regression) reported in Column 3 of Table 2 suggest that the model fits the data well, whichever estimation method we choose. Second, Figure 5a shows the log-log plot of the in-degree distribution fitted by the network formation model, in comparison with that from the data. It is clear that the model fits the data well, and the fit is much improved from that of the random network, as shown in Figure 5b.

To sum up the main findings, during 2008-2010, real estate agents rely substantially on network-based searches to find trading partners: 34%–55% of partners are found through network-based searches. In addition, on average, a real estate agent trades with one out of every 5 to 10 agents she meets. Finally, the model fits the data pretty well along several dimensions.

6 Robustness Analyses

In this section, we do two robustness analyses. First, we distinguish between experienced agents—who had more than 10 trades in the data period—and the other agents, and try to find out which type of agents relies relatively more on the network search. Second, one potential problem of the agent network we have studied is that it is incomplete in the sense that some agents may have joined the network before 2008 (the first year of the data period), but we do not observe their links formed before 2008. To solve this issue, we focus on the “new” agents who had no trade in 2008. These agents likely joined the network after 2008. Our data set contains a complete network of these agents as it includes all their links formed since they joined the network. In this section, we focus on the reiterative OLS estimation.
6.1 Network-Based Search is More Important for Experienced Agents than for Inexperienced Agents

We defined *experienced agents* as the top 10% real estate agents in terms of the number of trades; these are the agents who had more than 10 trades during the data period. The rest are *inexperienced* agents. The subsample of experienced agents has 154 agents (vs 1,588 agents in the full sample) who are involved in more than half of the trades (1,907 out of 3,643). Characteristics of the experienced and inexperienced agents are reported in Columns (2) and (3) of Table 1, respectively.

On average, each experienced agent has 24.77 trades and 21.13 trading partners, implying that the average number of trades per partner is 1.17. On the other hand, inexperienced agents have on average 2.42 trades and 2.34 partners only, with the average number of trades per partner being 1.035.\(^7\) This result implies that, compared to inexperienced agents, experienced agents have gained more trades by trading with more partners, instead of trading more frequently with the same partners.

Next, we reestimate the parameters, \(r\) and \(p\), for the network among experienced agents and the network among the inexperienced agents, respectively.\(^8\) The results are reported in Table 3. First, the fraction of transitive triples \(C^{TT}\) measures how likely we observe “partners of partners being partners” in a given network; A larger value of \(C^{TT}\) in general implies more links in the network are formed through network-based meetings. We observe that \(C^{TT}\) is much higher for experienced agents than for inexperienced agents (5.5∼10.1% vs 0.2%), suggesting that network-based search is more important for experienced agents than for inexperienced agent. Moreover, the ratio of random search to network-based search \((r)\) is much smaller for experienced agents than inexperienced agents (0.461∼1.087 vs 6.13∼1,998,486), which also implies that experienced agents rely more on the network-based search than inexperienced agents do. Note that an extremely large value of \(r\) implies that the network is

\(^7\) \(2.42/2.34=1.035\).

\(^8\) Note that in the process, we lose the links between experienced agents and inexperienced agents.
almost uniformly random. The results are robust to the various assumptions on who—the listing agent or the buyer agent—initiates the trades.
Table 3: Estimated Parameters for the Subsamples
(Experienced vs Inexperienced Agents)

<table>
<thead>
<tr>
<th>Assumptions on who initiated the trade:</th>
<th>Network of experienced agents</th>
<th>Network of inexperienced agents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>All trades are initiated by buyer agents</td>
<td>5.110</td>
<td>5.110</td>
</tr>
<tr>
<td>All trades are initiated by listing agents</td>
<td>0.055</td>
<td>0.101</td>
</tr>
<tr>
<td>Trades are equally initiated by listing &amp; buyer agents</td>
<td>0.461</td>
<td>0.882</td>
</tr>
<tr>
<td>All trades are initiated by buyer agents</td>
<td>0.412</td>
<td>0.970</td>
</tr>
<tr>
<td>All trades are initiated by listing agents</td>
<td>0.988</td>
<td>0.925</td>
</tr>
</tbody>
</table>

Note: $m$ is the average out-degree of the agents. $C^{TT}$ is the fraction of transitive triples as defined in (6). $r$ is the ratio of random search to network-based search. $p$ is the fraction of an agent’s friends who are also trading partners of the agent. $R^2$ is a measure of the goodness-of-fit of the model.
6.2 New Agents Rely Mainly on the Random Search

As we've mentioned earlier, the agent network that we observe in the full sample is incomplete in that we do not observe links that are formed before 2008. To address this issue, we focus on the “new” agents defined as those who had no trade in 2008 (the first year of the data period). These agents likely joined the trading network after 2008, and we observe the complete network among these “new” agents.

As shown in Column (4) of Table 1, there are 179 “new” agents with 102 links. The “new” agent has an average of 1.14 trades with 1.14 trading partners, so exactly one trade per partner. In Table 4, the estimated value of $r$ (3.046~43,903.610) suggests that “new” agents find partners mainly through random searches: 75%~100% of partners are found through random searches.\(^9\) Note that all “new” agents are inexperienced as they have less than 5 trades in the data, therefore this result coincides with our previous result that inexperienced agents rely more on the random search.

<table>
<thead>
<tr>
<th>Assumptions on who initiated the trade:</th>
<th>(1) All trades initiated by buyer agents</th>
<th>(2) All trades initiated by listing agents</th>
<th>(3) Trades are equally initiated by listing &amp; buyer agents</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>1.140</td>
<td>1.140</td>
<td>1.140</td>
</tr>
<tr>
<td>$C^{TT}$</td>
<td>0.000</td>
<td>0.000</td>
<td>N/A</td>
</tr>
<tr>
<td>$r$</td>
<td>3.046</td>
<td>N/A</td>
<td>43,903.610</td>
</tr>
<tr>
<td>$p$</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.965</td>
<td>N/A</td>
<td>0.966</td>
</tr>
</tbody>
</table>

Note: This table reports the estimated parameters for the network of “new” agents who had no trade in 2008. In Column (2), the regression does not converge due to lack of enough observations. $m$ is the average out-degree of the agents. $C^{TT}$ is the fraction of transitive triples as defined in (6). $r$ is the ratio of random search to network-based search. $p$ is the fraction of an agent’s friends who are also trading partners of the agent. $R^2$ is a measure of the goodness-of-fit of the model.

To sum up, the robustness analyses show that agents rely mainly on random searches to find trading partners at the early stage of career, and rely more on network-based searches

\(^9\) $3.046/(1+3.046)=75\%$ and $43,903.610/(1+43,903.610)=100\%$. 

20
after they have had many trades. One possible reason for this behavior pattern is that a new agent often has no local network to rely on, so she has to randomly meet other agents to start building her local network. After the new agent becomes seasoned and has a large local network, she starts to rely more on network-based searches.

7 Conclusion

This paper fits Jackson and Rogers' (2007) network formation model to the trading network of real estate agents. Fitting of the model generates three main results. First, in the process of finding a trading partner, real estate agents rely substantially on network-based searches during 2008-2010: 35%−55% of partners are found through network-based searches. Moreover, agents tend to randomly search for trading partners at the early stage of career, but rely more on network-based searches after they have had many trades. Second, during the data period, on average a real estate agent trades with one out of every 5 to 10 agents she meets through either random or network-based searches. Finally, the model fits the data pretty well along several dimensions.

A few interesting questions naturally arise from the current model. First, for analytical tractability, we assume the agent trading network is directed, and the network-based search occurs only along links leaving an agent’s friends. A more realistic assumption could be that agents conduct network-based searches through all links of their friends, and that network-based searches are conducted among friends-of-friends (instead of among partners-of-friends). As shown by Jackson and Rogers (2007), however, this alternative assumption makes the model much more complicated.

Second, there remain important questions on the formation of the network. For instance, is each link between agents formed by chance or by choice of the agents? How do geographic and brokerage factors affect the formation of the network? To answer these and related questions, we need a strategic network formation game theory model where the cost and the
benefit of forming and/or deleting each link is considered. This is out of the scope of the current paper but an important topic for future research.

Finally, it is interesting to fit the model to data from other geographic areas and/or other periods, to see if the results can be generalized. In particular, the market condition during 2008-2009 (right after the burst of the subprime mortgage crisis) is very different from other periods; sales are slow and properties are difficult to sell. If we think of random search as searching on the MLS and network-based search as searching within the same brokerage firm, Dale-Johnson and Hamilton (1998) have shown that brokers are more likely to do random search during slow market conditions. If this is the case, then the relative importance of the network-based search might have been underestimated in this paper.
8 References


Estate Economics, 20: 533-548.


Appendix

In the Appendix, we prove Prediction 1.

Proof. In this proof, we will derive from (4) the formula for the in-degree distribution $F_t(d)$ for a fixed $t$. First, from (4), we have:

$$d_i(t) < d_j(t) \quad \text{if and only if} \quad i > j. \quad (8)$$

Note that $i$ is the time when agent $i$ enters the network. Therefore, (8) means that the in-degree of a node $i$ is smaller than that of $j$ if and only if $i$ enters the network after $j$. Therefore, if we observe that a node entering at time $i$ has in-degree of $d$, then the fraction of nodes that have in-degree no more than $d$ is $(t - i)/t$. That is, $F_t(d) = 1 - i/t$.

Define $i(d)$ as such that $d_{i(d)}(t) = d$. That is, $i(d)$ is the node which has in-degree of $d$ at time $t$. Then from the above analysis, we have:

$$F_t(d) = 1 - \frac{i(d)}{t}. \quad (9)$$

Next, by substituting $i(d)$ into (4), we have:

$$d = (d_0 + rm) \left( \frac{t}{i(d)} \right)^{1+r} - rm. \quad (10)$$

Rearranging (10) gives:

$$\frac{i(d)}{t} = \left( \frac{d_0 + rm}{d + rm} \right)^{1+r}. \quad (11)$$

From (9) and (11), we have:

$$F_t(d) = 1 - \left( \frac{d_0 + rm}{d + rm} \right)^{1+r}, \quad (12)$$

which is exactly (5). This completes the proof. \qed