SMM638

Group 7

CASS Business School

Network Analytics Companion Document

## **Introduction**

## During our Network analysis, we set out to explore the world of football and the effect of homophily on player performance. We defined homophily in a membership network where *k* equals the number of previous teammates a player has on the club that he is joining. This two-mode network is made up of players and clubs as vertices and their transfer activities as edges. Our goal was to investigate the existence of homophily and how it affects a player’s future transfer fee and his tenure. This homophily exists in the form of membership closure where one player is likely to join a new club where he has previous teammates who play for that club. This relationship is represented by *p(k)*, the probability of joining a new team where *k* alters exist. We conducted the hypothesis test as followed to observe any possible relationship homophily has with future transfer fee and tenure:

## **Homophily:**

*Null hypothesis: Homophily has no effect on transfer-decision making*

*Alt hypothesis: Homophily has positive effect on transfer-decision making*

**Performance:**

*Null hypothesis: Homophily has no relationship with future fee*

*Alt hypothesis: Homophily has a relationship with future fee*

**Tenure:**

*Null hypothesis: Homophily has no relationship with tenure*

*Alt hypothesis: Homophily has a relationship with tenure*

## **Tests and Methods**

In this section, we tested the homophily hypothesis and worked on the correlations homophily may have with future transfer fee and tenure individually.

### **2.1 Homophily**

#### **2.1.1 Membership Closure analysis**

In this football network, we observe homophily after an event when a node (player) joins a new club where *k* number of previous teammates are on the roster. This type of closure resembles membership closure where the connection with a previous teammate influences a player’s decision to join that specific team or the manager's decision to acquire that specific player. Our primary analysis shows a decreasing *p(k)* as the number of *k* increases (**Figure 1**). However, the existence of homophily cannot be deduced by the relationship between *k* and *p(k)*.

*A picture containing shape

Description automatically generated*

**Figure 1.** **Relationship between *k* and *p(k)* in the football network.**

### **2.1.2 Test the existence of Homophily in the Football Network**

In order to conclude whether homophily exists in the network, we simulated 10 random networks resembling the football data (**Figure 2**). The simulated data closely followed the proportion of the number of transfers by players in the real network. Additionally, the simulated data assumes players only transfer within the nine major leagues and are always in a club until 2020. We do not include youth teams or retired players due to the limitation of the data.

*Graphical user interface, calendar

Description automatically generated*

**Figure 2. Comparison between** **our network and randomly simulated network**. Each subplot demonstrates the difference between *p(k)* of the real data and the average *p(k)* of the simulated data along with the z-score at a specific *k*.

From **Figure 2**, we can confidently reject the null hypothesis that homophily does not exist in our network.

### **2.2 Homophily’s effect on a Player's Performance**

#### **2.2.1 Homophily’s effect on Transfer Fee**

At the beginning of the performance test, we noticed there seemed to be a trend between age and fee. The fee movement in relative to age is shown in **Figure 3**.

A picture containing icon

Description automatically generated

**Figure 3. Bivariate distribution between age and average fee.**

In order to account for the correlation between these two variables, we decided to use an adjustment coefficient.

After observing the distribution of the data (**Figure 4**), we noticed a heavy tailed distribution and concluded on using the median as a measure of central tendency for our adjustment coefficient due to its resilience against extreme values.

Shape, square

Description automatically generated

**Figure 4**. **Histogram of fee**

Based on the distribution of the median fee and age, we utilized second-degree polynomial to smooth the relationship (**Figure 5**). An increase in fee is observed until the peak at the age of 25 where the trend reverts.

Chart

Description automatically generated

**Figure 5**. **Distribution of fee changed over age.**

We generated age-adjustment coefficients by using the reciprocal of the polynomial smoothed coefficients. The coefficients for ages with less than 100 observations were set to their closest coefficients as they deemed to be most similar (**Figure 6**).

Shape

Description automatically generated

**Figure 6. Age-Adjustment coefficient distribution**

After making this adjustment we went ahead to test the hypothesis. We only considered players who had two or more transfers because *k* depends on previous teammates. The performance measure we are considering is the increment in transfer fee a player is worth after playing for a team. We decided to remove all transfers that were not fee transfers (i.e. loans, free transfers, etc.) because they are not an appropriate proxy for performance. The result demonstrated a negative relationship between the change in fee and age, and thus the null hypothesis is rejected (**Figure 7**).

Chart, line chart

Description automatically generated

**Figure 7. Line chart between *k* and change in fee**

#### **2.3 Homophily’s effect on Tenure**

Next, we performed a tenure analysis to check our hypothesis on tenure. After performing an analysis on the relationship between tenure and homophily, we conclude that homophily negatively affects how long a player will stay on a team. **Figure 8** demonstrates that tenure is negatively correlated with homophily and thus we reject the null hypothesis.

Chart, line chart

Description automatically generated

**Figure 8. Line chart between *k* and number of years stayed**

## **Discussion on Ibrahimovic’s Example**

To further elaborate on our analysis, we drew conclusions from a relevant example. When Zlatan Ibrahimovic moved from Barcelona to AC Milan (he recently left Inter Milan to go to Barcelona) he moved to a team where *k* was very high. Based on reporters who covered the event, Zlatan moved because he was unhappy with his former manager Pep Guardiola and he had many old teammates currently playing for AC Milan (2010, The Guardian). In other words, his move was heavily based on relationships. A year after this transfer AC Milan realized Zlatan was not the player they thought he was and he did not fit on their current roster. He was quickly sold to Paris for a lower fee than they initially bought him for. Essentially, when a player or manager makes a transfer decision based on homophily, our data tells us that something (i.e., unrealistic expectations, too much pressure on relationships, or lack of thought put on the football aspects of the movement) caused the player to perform at a lower level and therefore warrant less money in his next transfer. Additionally, we observe that in the high homophily scenario the player also spends just over 12 months on AC Milan, which supports our second assumption of low tenure with high homophily. If a player joins a team for a short amount of time, one would expect there were some negative circumstances overshadowing his stay on that specific team. If there is negativity, either in his play or relationships with teammates/manager, the player would not be worth as much on the market. Therefore, homophily negatively affects performance.

After running all these analyses, we can conclude that Homophily exists; a negative relationship between homophily and transfer fee exists, along with a negative relationship between homophily and tenure. To dive into this concept even further, we decided to try to find some examples using comparably skilled players. Using a FIFA dataset and overall FIFA rating as the criteria, we decided to only include above-average players in this analysis (Z-score above). We wanted to see how the results of these players may differ from the overall football data in terms of their relationships with homophily, future fee, and tenure. We thought this would be interesting because using the data of thousands of players gives a good representation of the sport of football as a whole. However, most people watch to see their favourite players, hence we wanted to conduct a study on the outliers to see if they have the same trend as the norm. **Figure 9** shows that the trend for the competent players including outliers is somewhat similar to the trend for the entire dataset. The few differences lie in the very high homophily scenarios where fee actually increases at certain points. This could be because age increases opportunities to join a high k situation are naturally higher for players. Elite players tend to perform better at older ages than average players and therefore their fee might decrease slower as they age (and in turn enter high k situations) than the average footballer.

|  |  |
| --- | --- |
| Chart, line chart  Description automatically generated | Chart, line chart  Description automatically generated |

**Figure 9. Line chart between *k* and change in fee / year stayed of top players**

Furthermore, Ibrahimović was playing for Barcelona and was having issues with the manager Pep Guardiola (2010, The Guardian). He began looking around for a new situation and because he recently left Inter Milan to join Barcelona, he naturally looked to the place he was most familiar, Milan. AC Milan came calling and the combination of vengeance for his old club and familiarity with the players on AC Milan led him to this new club. “I am familiar with Alexandre Pato and Ronaldinho, we have a fantastic attack. Now it depends on us. I think Milan fans will enjoy watching the three of us play at the stadium.” Along with the two players mentioned above Ibrahimović had many connections to AC Milan and it seems his decision was mostly made based on his familiarity with the squad. After a bitter great reunion and glorious reception by the AC Milan players and faithful, the situation couldn’t have been worse. Ibrahimovic was not the player many of the AC Milan players remembered and he did not truly fit the team in Milan. Just 12 months into his contract AC Milan shipped him out to Paris Saint Germain for less money than they bought him for 12 months prior. “He has always been the one to decide his destiny, except on this occasion” said the president of AC Milan years later.

## Reference

The Guardian, (2010), *“Milan agree deal to sign Zlatan Ibrahimović from Barcelona”*, [online], Available at: <https://www.theguardian.com/football/2010/aug/28/zlatan-ibrahimovic-milan-barcelona> (Accessed 15 December 2020)