Personality Correlation Analysis and Applications in Social Networks

Te-Min Chang
Information Management, National Sun Yat-sen University, Kaohsiung, temin@mail.nsysu.edu.tw

Follow this and additional works at: http://aisel.aisnet.org/amcis2012

Recommended Citation
http://aisel.aisnet.org/amcis2012/proceedings/EBusiness/5

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2012 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Personality Correlation Analysis and Applications in Social Networks

Te-Min Chang
Department of Information Management
National Sun Yat-sen University
temin@mail.nsysu.edu.tw

ABSTRACT
The focus of this study is to investigate the personality similarity among people. We investigate 203 relevant personality adjectives to determine their stable intrinsic values. In addition, the contextual connotation of every adjective is introduced to measure personality correlation. It is then applied to several social network groups to help interpret the group development. The results show that (1) in naturally formed interpersonal groups, openness to experience is the most important formative indicator, (2) groups under high task environment seek to take conscientiousness as the primary formative indicator to converge into task groups, (3) a lovers’ relationship assumes partial personality similarity as the predominant phenomenon, and (4) The convergence of group personality shows comparatively obvious concentration.

KEYWORDS
Big Five model, social network, personality similarity.

INTRODUCTION
Human resource is a vital factor of production in a particular organization, and is critical to output quality and efficiency. Since Hawthorne’s study in the 1930s, the academic community of sociology has increasingly accepted research perspectives involving “social man” instead of “economic man”, which opens the field of behavioral psychology. According to surveys, 68 % of the top 1,000 Fortune enterprises adopt teamwork design. Because the body of a team is people, the interaction among people is essential in teams. The effect of interaction profoundly affects team performance.

Group development can be attributed to human personalities. However, currently only abstract concepts are applied to address personality similarity without quantifying ones. How to quantitatively define the degree of similar personalities has become a considerable challenge in related studies. A specific comparison begins with the description of adjectives ascribed to two persons. Nonetheless, recent research has not been able to determine a specific number of adjectives that can be used to describe human personalities. If such a work can be done, the degree of similarity of personalities between groups can then be discussed. This study is thus based on the concept of establishing personality correlation analysis, with the hope of revealing groups with similar personalities in interpersonal relationships, while enhancing the function of interpersonal relationships and promoting social development.

This study adopts personality adjectives in language as a medium to complete the personality coding of language passwords. The major research purposes include: (1) establishing an approach to analyzing personality correlations using the five-factor model of personality, (2) developing a quantitative formula to determine similarity among adjectives (that is, the degree of similarity matrix); and (3) determining the characteristics of interpersonal patterns and the impact of similarity variables employing different types of group verifications.

RESEARCH FRAMEWORK
The research framework of this study contains three parts (Figure 1). The first part is the entire organization of database establishment. Because the method involves matching personalities with language adjectives, an independent step toward establishment from screening adjectives is required. The second part entails analyzing data in the database, the major purpose being the approach of analysis and specific programming. Calculating vectors, distinguishing levels of strength, designing logic, programming code composition, and completing an analytic system form another independent step. The third part involves operating and applying the analytic system in different social network groups to observe composition characteristics and phenomena, as well as engaging in primary exploration and discussing results.
SCREENING OF ADJECTIVES

During the process of literature review, this study carefully collected every uni-polar adjective (having a significant single score) used by academia in the past to establish a preliminary data list composed of 235 words, which were condensed to 203 words after significant amendments. The code sorting uses five letters: E, A, C, N and O in the Big Five model (Digman, 1990; Ekehammar and Akrami, 2007, Rothmann and Coetzer, 2003; Erdheim et al., 2006; Argamon et al., 2009). For example, "A03 Virtuous" means that it is a Type A adjective and is the third being included in the list. Because every adjective was defined as holding an independent personality score during the collection process, A03 Virtuous is actually defined as having the personality of agreeableness; thus, it was sorted into class A03 in the order of collection. The number of adjectives collected according to each code list is indicated in Table 1.

<table>
<thead>
<tr>
<th>Personality traits</th>
<th>Extraversion</th>
<th>Agreeableness</th>
<th>Conscientiousness</th>
<th>Neuroticism</th>
<th>Openness to Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial code</td>
<td>E</td>
<td>A</td>
<td>C</td>
<td>N</td>
<td>O</td>
</tr>
<tr>
<td>Number collected</td>
<td>44</td>
<td>37</td>
<td>47</td>
<td>39</td>
<td>36</td>
</tr>
</tbody>
</table>

Table 1. Statistics of Personality Trait Adjectives Collected

Although equal collection of every trait was pursued in the collecting process, the collection quantities of extraversion and conscientiousness are more than others, which may be related to diversified expression in Chinese. In contrast, the collection quantity of openness to experience is the fewest, as there are less variant related adjectives in Chinese.

QUESTIONNAIRE DATA COLLECTION

In the action of matching every target adjective with its personality, this study applies the questionnaire approach to provide a set of Big Five personality scores. Figure 2 shows the core model example of questionnaires. Each target adjective possesses an exclusive scale with five scores, which represents the measure of the Big Five scores including extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience (from top to bottom). Starting from neutral 5, the maximum to the right is 10, meaning the strongest score; the minimum to the left is 0, meaning the weakest score. In the process of the questionnaire survey, respondents were invited to measure how the personality Big Five scores of the target adjective are distributed.

The questionnaire applies the design of open dimension, enabling respondents to select randomly the degree of interval and then capture the value required. Every questionnaire measures seven adjectives; thus, there are seven groups of scales. Moreover, because 203 adjectives must be measured, 29 versions are designed and arranged randomly, enabling adjectives to be distributed evenly in every version. When releasing mass questionnaires, balanced distribution and non-redundant placement is conducted to release 29 versions simultaneously and enable equal probability for every version to avoid
unbalanced distribution of samples. Because the credibility can be maintained by achievement of 30 samples for normal distribution approximation, every version must have at least 30 valid responses. This means that every adjective must be measured more than 30 times to gain credibility. Thus, 40 copies of every version are prepared when releasing questionnaires. With the number of versions being 29, 1160 (29 × 40) copies are released. The test subjects are students and faculty members mainly from a university using a convenience sampling method. Of the 1160 questionnaires released, 1001 were returned, among which 28 are incomplete. Therefore, there are totally 973 valid questionnaires and the valid response rate is 83.87 %.

ESTABLISHMENT OF ADJECTIVE DATABASE

After releasing questionnaires, the results of 203 adjectives were analyzed to acquire an internal personality score and establish an adjective database. The major concept involved assigning every adjective a group of personality scores, for the convenience of vector calculation with other adjectives to be compared later. Simple values were provided for a dot-and-dash line on spot charts. The line in the middle represents a neutrality of 5; exclusive personalities belonging to this adjective can be seen in the charts (see Figure 3 for example).
Figure 3 shows the analysis of the adjective "C33 Cautious" where every numerical point representing opinions of respondents. Each respondent assigned five gradings: horizontal E, A, C, N, and O. Each line from top to bottom represents a different respondent and there are 40 respondents in this case. The red middle line is the location of mean value. Every adjective is evaluated by 34 respondents on average. In addition, Table 2 shows the summary statistics of the respondents’ evaluations toward "C33 Cautious." The decision value is designed to convert the degree interval between 0 and 10 into an interval between −5 and +5. Thus, in terms of observing decision values, those scores whose absolute values are greater than 1 can be regarded as having such a trait score and are expressed in the word "possess". In the subdivision of those possessing such traits, we use the term "significant score" to define those with absolute values between 1 and 3; we use the term "strong score" to define those with absolute values between 3 and 5. As shown in the Table 3, this adjective possesses one strong score of C and two significant scores of E and N traits.

<table>
<thead>
<tr>
<th>Decision value</th>
<th>E Score</th>
<th>A Score</th>
<th>C Score</th>
<th>N Score</th>
<th>O Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>−1.16</td>
<td>−0.35</td>
<td>3.29</td>
<td>1.39</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Original value</td>
<td>3.84</td>
<td>4.65</td>
<td>8.29</td>
<td>6.39</td>
<td>5.13</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1.474</td>
<td>1.417</td>
<td>1.295</td>
<td>2.030</td>
<td>1.832</td>
</tr>
<tr>
<td>Upper bound of error</td>
<td>4.033</td>
<td>4.842</td>
<td>8.469</td>
<td>6.664</td>
<td>5.375</td>
</tr>
<tr>
<td>Length of Interval</td>
<td>0.390</td>
<td>0.390</td>
<td>0.350</td>
<td>0.548</td>
<td>0.485</td>
</tr>
</tbody>
</table>

Table 2. Example of Data Statistics of Adjective “C33 Cautious”

**ANALYSIS FOR ADJECTIVE CORRELATION MATRIX**

After establishing the adjective database, this study designs a personality correlation degree matrix, which is established from the comparison of the values calculated from the 203 target adjectives. The matrix thus consists of 41,209 values, each of which representing a degree of personality similarity.

Since every adjective is expressed by decision values in terms of the Big Five dimensions, it is intuitive to compute the Pearson’s correlation or cosine-based similarity between any two adjectives. However, recall that the decision value of 0 indicates null meaning coverage of the specific personality dimension. Therefore, two adjectives with scores of 0 on the same dimension do not indicate being similar; instead, it only means that both are not related to that specific personality dimension. Simply employing Pearson’s correlation or cosine-based similarity will introduce misinterpretation among adjective similarity.

To amend such a problem, this study proposes a computing formula for the adjective correlations as shown in eq. (1):

\[
\sum_{i=E,A,C,N,O} \gamma_i \times \left\{ 1 - \frac{|X_i - Y_i|}{10} \right\}
\]

where \(X_i\) and \(Y_i\) are the decision values of two adjectives on the \(i^{th}\) dimension, and \(\gamma_i\) is a strength weight on the corresponding dimension to offset the impact resulted from closeness (not similarity) of null meanings.

The formula of \(\gamma_i\) weight is defined as follows:

\[
\gamma_i = \frac{\sum_{i=E,A,C,N,O} |X_i + Y_i|}{2 \sum_{i=E,A,C,N,O} |X_i| + |Y_i|}
\]

eq. (2)
With eq. (2), if both decision values of two adjectives are close to 0 on the same dimension, the corresponding weight will approach 0 and thus diminish its original effect of being falsely recognized as similar.

For example, suppose that the decision values for two adjectives are (1, 2.5, 1, 1, 1) and (1, 3.5, 1, 1, 1), respectively. That is, both have strong indication on the A dimension but relatively irrelevant to other dimensions. The strength weight \( \gamma_A \) is computed as

\[
\gamma_A = \frac{(2.5 + 3.5)}{(2.5 + 3.5) + \sum_{i=E,C,N,O} (1 + 1)} = \frac{6}{6 + 4 \times 2} = 0.4286
\]

while the weights for all other dimensions are the same as 0.1429. The similarity between these two adjectives is therefore 0.3857 while the results for Pearson’s correlation and cosine-based similarity are as high as 1 and 0.6955, respectively. The proposed formula effectively offset the impact of E, C, N, and O dimension scores as the two adjectives are not related to those dimensions instead of being similar.

After applying the above computation to all paired adjectives, we are able to obtain an adjective correlation matrix. Table 3 displays the concept of the adjective correlation matrix. For example, the similarity between "A01 Trustworthy" and "A08 Receptive" is merely 44%, meaning that their internal scores are different with low similarity, though both of them belong to "Type A" (adjectives of agreeableness). To distinguish easily, similarities lower than 50% are displayed in red.

<table>
<thead>
<tr>
<th>Code</th>
<th>A01</th>
<th>A02</th>
<th>A03</th>
<th>A04</th>
<th>A05</th>
<th>A06</th>
<th>A07</th>
<th>A08</th>
</tr>
</thead>
<tbody>
<tr>
<td>A01</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A02</td>
<td>76%</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A03</td>
<td>63%</td>
<td>71%</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A04</td>
<td>61%</td>
<td>76%</td>
<td>79%</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A05</td>
<td>68%</td>
<td>95%</td>
<td>78%</td>
<td>95%</td>
<td>100%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A06</td>
<td>63%</td>
<td>71%</td>
<td>98%</td>
<td>73%</td>
<td>66%</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A07</td>
<td>57%</td>
<td>78%</td>
<td>78%</td>
<td>96%</td>
<td>92%</td>
<td>69%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>A08</td>
<td>44%</td>
<td>59%</td>
<td>85%</td>
<td>48%</td>
<td>57%</td>
<td>91%</td>
<td>48%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Table 3 Example of Adjective Correlation Matrix

**ANALYSIS OF PERSONALITY SIMILARITY**

Furthermore, this study discusses the "relative" relationship among interpersonal network; that is, the degree of similarity among personalities, rather than an "absolute" factor of an individual. The purpose is not to discuss the absolute values of the five factors of an individual's personality, but instead the relative values among them. The major focus lies in the degree of similarity, but not in which type of personality an individual has. This study places emphasis on the "relationship" within a group, with aim to develop an approach to discussing and exploring the special characteristics and reasons for group composition regarding interpersonal phenomena.

Considering that reading excessive adjectives confuses respondents and consumes a considerable amount of time, this study hypothesizes that choosing 5 adjectives by any respondent is optimal. Assume that A and B in Figure 4 refer to two respondents. We can then obtain 25 correlation values from the adjective correlation matrix for their chosen adjectives. The average of these 25 values, 34.96%, indicates the personality similarity between A and B. It can be determined from the distribution of adjective distribution that A seems to be energetic and extroversion but a little over-ambitious; while B is relatively conservative, introversion, cautious, and thoughtful. The differences between these two persons' personalities therefore result in the personality similarity of 34.96%.
**IMPLEMENTATION AND FINDINGS**

With the aforementioned methods, this study establishes the concept of personality similarity analysis and designs a practical analytic system, as indicated in Figure 5. The overall personality similarity and individual Big Five correlation degrees are listed as major outputs. Figure 5 shows that A and B, two actual respondents, are highly correlated in their personalities of the C score, conscientiousness, through the explanation of personality similarity by this system, which reflects that A and B are similar in their viewing of things and the cognition of responsibilities. In addition, we set up a threshold of similarity degree over 50% as high by observing from 70 respondents personality similarity values that ranges between 40% and 60%. Thus, we take the middle value as the basis of comparison as a rule of thumb. Certainly, minute differences among values can also be used for interpretation. For example, a similarity degree of 58% is apparently higher than 52% and 53%.

Finally, this study attempts to apply the personality similarity results in various social network group styles for verification and discussion. The social network group styles include lovers’ relationships, a community group and two interpersonal groups. The lovers’ relationships exhibit the simplest form. Only parallel lines among lover pairs are analyzed to arrive at clear observation results. In contrast, the observation of the community group and the interpersonal groups requires longer identification time to avoid temporary phenomena. Finally, the key to distinguish the community group and the interpersonal group lies in “whether the group is established on tasks or not.”
LOVERS’ RELATIONSHIP

Lovers’ relationship is the simplest group relationship, which is composed of single pairs. This study collected 10 groups of lovers for examination. Figure 6 shows the results. We observe that none of the 10 groups have significant personality similarity since all RS’s (Relationship Similarity) are below 50%. However, seven groups form at least one line in the Big Five personality similarity, showing that these lovers’ personalities are not similar though they have something in common. The common traits vary in different pairs of lovers. Some are similar in agreeableness while others are similar in neural sensitization, thus forming a diversified lovers’ relationship. Thus, a lovers’ relationship is not a combination completely sharing the same ambitions and purposes; instead, it is a mutual complementation in personality with a small number of common traits.
COMMUNITY GROUP

To apply the methodology proposed in this study, the researcher observed a community group (denoted as LD01) formed by 10 members who are leaders of music clubs at the university. This group is a task-based group with low personality similarity, a high performance characteristic, and diversified members while it has led itself to win the third place in a national music competition and the perfect level in a national club evaluation. Initially, this group was formed due to task requests but finally became a group with friendship characteristics. The members belong to ten different departments, some of them having maintained frequent communication after the tasks were completed, strengthening the friendship core, and holding full member parties frequently.

Results in Figure 7 show that LD01 has the characteristic of $\text{GS (Group Similarity)} = 33\%$, which is a relatively low personality similarity. However, regarding of a group with 10 people, its complexity is higher because several subgroups can be separated with an easily conducted additional screening. Due to four-year's time being together, different degrees of a close relationship have been developed among members under the interaction theory; and thus the separation discussion of subgroups within the group requires other social network instruments to complement. This LD01 is a group of club leaders with a boundary line, but more friendship groups are formed over time and make the content of group relationships rich.

Moreover, if this study reorganizes a relationship chart with six members including 3, 4, 7, 8, 9, and 10, a new core will be formed with extremely high personality similarity, holding the condition of becoming a friendship group. However, such a scenario does not reflect reality. This new core is not the most intimate; instead, it exists in dispersed small groups. This indicates that, in large groups, the personality similarity within the group is not the only indicator for group development. Subgroups with respective to friendship are formed (not completely complying with the theory of similarity), but with other influencing factors; for example, demand theory and sociality affect the formation of friendship groups, as presented by LD01.
INTERPERSONAL GROUPS

This study designs a test for interpersonal groups with high and low similarities where FA01 is the verification group with high similarity (See Figure 8 for the results). This group is a small group in class, taking actions and participating in many reports together during the college four years. This group is a highly essential group in the sampling of this study. Showing an extremely high personality correlation (GS = 83%), five of the six relationship lines display personality similarity and a convergent personality of Type O, indicating that this group has the same temperament regarding opinion and curiosity of new things. This is also because the group has characteristics of "friendship and task" simultaneously. The members take actions together and complete tasks together. They do not leave the action because the task is finished; instead, they remain in the group. It is a typical group with a high degree of coincidence. This study also infers that these group members and friends sharing the same ambitions and purposes attract each other because of the similarity of Type O scores.

In contrast, FB01 is a verification group with low similarity (See Figure 9 for results). This group is a fixed interpersonal circle with eight years of being together. It developed from the relationship of classmates in primary school and has performed many activities such as budget traveling. Their contact is still frequent and active thus far. However, the analysis results show that FB01 has an overall GS of 33.3%, which indicates a low group personality similarity. Nor do the group members share similar personality on any of the Big Five dimensions. It seems therefore in this case similar personalities do not help interpret relationship binding and keeping. Nonetheless, with respect to this long-term relationship group, one may conjecture that personality similarity may still help relationship building at the early stage; it is the "time" factor that allows members to develop different personalities while their primary relationships established long time ago remains intact.
CONCLUSIONS

This study is based on "similarity theory" to examine social network group development, and concludes that, in the process of system operations, similarity theory cannot completely interpret the formation of interpersonal relationships. This study also discovered in the process of discussing the possibility of "similarity" that openness to experience is the most critical formative indicator and the most principal reference frame of interpersonal groups. Different from the selection of interpersonal friends, this study also discovered that, if a group is formed under a high task environment, openness to experience can no longer be used as a criterion for selection when converging task groups because it might only represent the cost of communication. A task team might also require the precision and performance indicator of accomplishing the task; thus, the requirement of conscientiousness gradually becomes consistent. The intention to seek partners with identical conscientiousness and care when forming task groups to perform tasks and actions obviously exists. Reflecting on activities in classes and organizations, once a special task or goal arises, the method of selecting friends might begin to change. The primary indicator perhaps shifts from the focus on communication to that of the accomplishment of task and goal. Seeking people with identical conscientiousness, precision, and care similar to a particular person or organization comprises the most critical indicator. The shifting phenomenon also changes simultaneously.

Furthermore, this study verifies that a lovers’ relationship does not ask for a composition with complete consistency. Only a slightly sufficient degree of personality similarity is needed to bridge between the two parties to develop their relationships. For examples, lovers with similarity in openness to experience might think about the cost and method of communication; lovers with similarity in conscientiousness might focus on the consensus of responsibility cognition; and the similarity in neuroticism might indicate a certain type of special privacy and agreeableness that enables the two parties to face norms together. The major contributions of this work are to provide a different perspective for measuring similarities more reliably with validity to assist organizations in managing human resources.

REFERENCES


