Business Profiles and Personality Traits towards Social Network Team Building Processes

Orestes Evangelinos
Aristotle University of Thessaloniki
Faculty of Sciences
School of Informatics

Business Profiles and Personality Traits
towards Social Network Team Building Processes

Orestes Evangelinos

Supervisor:
Athena Vakali
Associate Professor

Thessaloniki, September 2013
Συνδυασμός Εταιρικού Προφίλ
με Χαρακτηριστικά Προσωπικότητας
για Συγκρότηση Ομάδων σε Κοινωνικά Δίκτυα

Ορέστης Ευαγγελινός

Επιβλέπουσα:
Αθηνά Βακάλη
Αναπληρώτρια Καθηγήτρια

Θεσσαλονίκη, Σεπτέμβριος 2013
Disclaimer

During the drafting of this dissertation, and as a part and product of it, an application has been created. This application draws and uses data from users’ LinkedIn accounts. In accordance with the August 5th 2013 revision of the LinkedIn API Terms of Use¹, before any such event, the user is explicitly informed of the facts and has to give permission in order for the data to be collected (Section III.B.4). These data are neither distributed nor made public to third parties and are solely intended for academic use. As such, each user has access strictly to his or her profile information (Section III.A.1). For the few examples presented in the dissertation, user anonymity is applied.

Abstract

Nowadays, businesses aim to produce products and offer services at an ever increasing rate. One way for them to achieve this is through redefined team building processes, where the teams created will be more cohesive and efficient. At the same time, the use of Social Networks has become, in recent years, quite widespread. LinkedIn is a business-oriented social network, where users have their business profile available online. Furthermore, there has been some headway in the research of team efficiency and performance, based on team personality composition. We are, therefore, suggesting a new team building method, which combines the users’ business profile, along with their personality profile. The former is drawn from each user’s LinkedIn account, while for the latter we offer a personality questionnaire, for each user to complete. The questionnaire is based on the Big Five Factor Model, which defines personality as a combination of five aspects: Openness to Experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism. Having these data, the variables of a team can be defined (team size, skills required for the task, desired personality traits, etc). Users who match these criteria are fetched from our database. We then proceed to create all possible teams, find the similarity between each team’s members and rank them according to this value. We have devised a similarity formula, which takes into consideration six distinct criteria: Skills, languages spoken, industries in which uses may belong to, geographic location, educational background and social connectivity. The proposed team has, accordingly, the greatest similarity value.
Περιλήψη

Στις μέρες μας, οι επιχειρήσεις στοχεύουν στην παραγωγή προϊόντων και στην προσφορά υπηρεσιών με έναν όλο και αυξανόμενο ρυθμό. Ένας πιθανός τρόπος για να το καταφέρουν αυτό είναι με την αναθεώρηση των μεθόδων που χρησιμοποιούν για τη δημιουργία ομάδων εργασίας. Με αυτόν τον τρόπο, οι ομάδες που δημιουργούνται θα έχουν μεγαλύτερη συνοχή και θα είναι περίσσοτερο αποτελεσματικές. Παράλληλα, τα κοινωνικά δίκτυα είναι ευρέως διαδεδομένα και η χρήση τους τα τελευταία χρόνια βαίνει αυξανόμενη. Το LinkedIn είναι ένα κοινωνικό δίκτυο με επαγγελματική κατεύθυνση, στο οποίο οι χρήστες διατηρούν το επαγγελματικό τους προφίλ online. Επίσης, έχει παρατηρηθεί πρόσδοκα στην έρευνα για την αποτελεσματικότητα και την απόδοση των ομάδων, υπό το πρίσμα της σύνθεσης της προσωπικότητας των μελών που την αποτελούν. Για αυτόν τον λόγο, προτείνουμε μία νέα μέθοδο δημιουργίας ομάδων, η οποία συνυπάρχει το επαγγελματικό προφίλ των χρηστών, καθώς και τα στοιχεία της προσωπικότητάς τους.

Το επαγγελματικό προφίλ ενός χρήστη σχηματίζεται αντλώντας πληροφορίες από το προφίλ του στο LinkedIn, αφού πρώτα μας δοθεί ρητά η έγκρισή του για αυτή την ενέργεια. Οι πληροφορίες που μπορούμε να αντλήσουμε είναι οι δεξιότητες και οι γνώσεις που έχει ένας χρήστης, οι γλώσσες που μιλάει, ο επαγγελματικός κλάδος στον οποίο ανήκει ή έχει δουλέψει στο παρελθόν, η τοποθεσία του και το ακαδημαϊκό του υπόβαθρο. Λαμβάνουμε επίσης πληροφορίες σχετικά με τον τρόπο με τον οποίο συνδέονται οι χρήστες μεταξύ τους στο δίκτυο του LinkedIn.

Για την εκτίμηση της προσωπικότητάς, προσφέρουμε ένα ερωτηματολόγιο, ώστε να το συμπληρώσει ο εκάστοτε χρήστης. Το
ερωτηματολόγιο αποτελείται από 50 ερωτήσεις και βασίζεται στο Μοντέλο των Πέντε Παραγόντων της Προσωπικότητας (Big Five Factor Model), σύμφωνα με το οποίο η προσωπικότητα είναι συνδυασμός πέντε στοιχείων: Το να είναι κάποιος δεκτικός σε νέες εμπειρίες (Openness to Experience), η Ευσυνειδησία (Extraversion), η Εξωστρέφεια (Extraversion), η Συγκαταβατικότητα (Agreeableness) και ο Νευρωτισμός ή η Συγκαταβατική Σταθερότητα (Neuroticism, or Emotional Stability). Οι ερωτήσεις αντιστοιχούν σε συγκεκριμένους παράγοντες της προσωπικότητας, δηλαδή για κάθε έναν υπάρχουν 10 ερωτήσεις. Κάθε ερώτηση παρουσιάζει στον χρήστη μια κατάσταση, ένα συναίσθημα ή μια στάση προς ένα θέμα (όπως «Μου αρέσει η τέχνη», «Αισθάνομαι άνετα όταν βρίσκομαι ανάμεσα σε ανθρώπους») και υπάρχουν 5 πιθανές απαντήσεις: Διαφωνώ Πολύ, Διαφωνώ, Είμαι Ουδέτερος, Συμφωνώ, Συμφωνώ Πολύ.

Για τη δημιουργία μιας ομάδας προσφέρουμε μια σειρά από κριτήρια αναζήτησης. Πρωταρχικό κριτήριο είναι ο αριθμός των μελών που θα απαρτίζουν την ομάδα. Τα υπόλοιπα αντιστοιχούν στις πληροφορίες που έχουμε αντλήσει από το LinkedIn και είναι οι δεξιότητές που πρέπει να έχουν τα μέλη της ομάδας, οι γλώσσες που μιλάνε, ο επαγγελματικός κλάδος που ανήκουν, η τοποθεσία τους (δηλαδή πόλη, χώρα ή ήπειρος), το ακαδημαϊκό τους υπόβαθρο και η κοινωνική σύνδεση που θα πρέπει να έχουν τα μέλη της ομάδας μεταξύ τους. Επίσης, υπάρχει το κριτήριο με το οποίο προσδιορίζεται το ψυχολογικό προφίλ της ομάδας, οριζόντας τιμές για τα πέντε χαρακτηριστικά της προσωπικότητας. Υπάρχει η δυνατότητα να οριστούν τιμές για ορισμένα από τα χαρακτηριστικά και τα υπόλοιπα να απενεργοποιηθούν και να μην ληφθούν υπ’ οίκον.

Στη συνέχεια γίνεται αναζήτηση στη βάση δεδομένων και επιλέγονται οι χρήστες που ταιριάζουν στα κριτήρια αναζήτησης που έχουν τεθεί. Έπειτα δημιουργούμε όλους τους πιθανούς συνδυασμούς των χρηστών αυτών σε ομάδες με το ζητούμενο μέγεθος. Στην περίπτωση που οι χρήστες είναι λιγότεροι από τον επιθυμητό αριθμό των μελών της ομάδας, παρουσιάζονται όλοι ως μία πιθανή και διαθέσιμη ομάδα.

Το επόμενο βήμα είναι να συγκρίνουμε τις ομάδες μεταξύ τους, για να δούμε ποια προβλέπεται να είναι η πιο αποτελεσματική. Έχουμε
σχηματίσει, λοιπόν, έναν τύπο, που υπολογίζει την ομοιότητα ανάμεσα στα μέλη της εκάστοτε ομάδας, η οποία λαμβάνει υπ' όψιν έξι κριτήρια: Δεξιότητες του κάθε χρήστη, γλώσσες που μιλάει, τον επαγγελματικό κλάδο στον οποίο μπορεί να ανήκει, την τοποθεσία, το ακαδημαϊκό υπόβαθρο και την κοινωνική διασυνδέσμευση που έχει με τα άλλα μέλη της ομάδας. Όλες οι ομάδες ταξινομούνται και παρουσιάζονται σε φθίνουσα σειρά. Η προτεινόμενη ομάδα έχει, ακολούθως, την μεγαλύτερη τιμή ομοιότητας.

Μέσα από αυτή τη διαδικασία και ως αποτέλεσμα της μεθόδου που προτείνουμε, δημιουργήσαμε μια εφαρμογή με το όνομα PROTEAS (PROfessional TEAms in Social networks – Επαγγελματικές Ομάδες σε Κοινωνικά Δίκτυα). Κάναμε ένα κάλεσμα για συμμετοχή χρηστών στην εφαρμογή και για να επιβεβαιώσουμε την ορθή λειτουργία της δημιουργήσαμε μια σειρά από σενάρια. Κάποια από τα προβλήματα που αντιμετωπίσαμε είναι η μικρή συμμετοχή, η ποιότητα των δεδομένων που συλλέξαμε, που ήταν πολλές φορές ελλιπή, και η μη συμπλήρωση του ερωτηματολόγου σχετικά με την προσωπικότητα.

Στο εγγύς μέλλον, θα θέλαμε να πραγματοποιήσουμε ορισμένα πειράματα μεγαλύτερου μεγέθους και έκτασης. Ένα κατάλληλο περιβάλλον περιβάλλον είναι ένα ακαδημαϊκό μάθημα, από το οποίο μπορούμε να αντλήσουμε πληροφορίες σχετικά με τους συμμετέχοντες και τις εργασίες που ολοκληρώνουν, ώστε να γίνει σύγκριση με τα αντίστοιχα αποτελέσματα της εφαρμογής και να εξάγουμε έτσι χρήσιμα συμπεράσματα.
Acknowledgements

First and foremost, I would like to especially thank my supervisor, Associate Professor Athena Vakali, for the chance to work with her and her inspired guidance.

In addition, a great thank you to Christos Zigkolis, who has been very cooperative and has been giving me invaluable advice for the past year.

And finally, many thanks to my friends and family: Yannis, Dora, Nefeli, Danae, Fotis and Ksanthi.
Index

List of Figures ............................................................................................................. xiii
List of Tables ................................................................................................................ xv

1 Introduction ............................................................................................................. 1
   1.1. Open Problems ................................................................................................ 3
   1.2. Dissertation Structure ..................................................................................... 4

2 Literature Review .................................................................................................... 7
   2.1. Personality and Individual Efficiency .............................................................. 7
   2.2. Personality and Team Efficiency ....................................................................... 8
   2.3. Other Factors Relating to Team Efficiency ..................................................... 9
   2.4. Literature Discussion ...................................................................................... 10

3 Basic Concepts ........................................................................................................ 11
   3.1. Professional Profile ....................................................................................... 11
   3.2. User Personality ............................................................................................. 12
   3.3. Professional Social Networking: The Case of LinkedIn ................................ 15
   3.4. Team and Team Forming ................................................................................ 17

4 PROTEAS: “Automated generation of PROfessional TEAms in Social networks” .......................................................... 19
   4.1. PROTEAS Methodology and Principles ......................................................... 19
   4.2. Application Front-End and User Data Collection ........................................... 21
   4.3. Personality Questionnaire ............................................................................. 22
   4.4. Team Building: Search Criteria ..................................................................... 25
   4.5. Similarity Measures ....................................................................................... 26
   4.6. Team Similarity and Ranking Algorithm ...................................................... 36
5 PROTEAS Implementation ............................................................................. 39
  5.1. Current User Interface ............................................................................. 39
  5.2. Team Building Search Interface ................................................................. 44
6 Experimentation and Discussion .................................................................. 49
  6.1. Search Scenarios ..................................................................................... 49
  6.2. Discussion ............................................................................................... 54
7 Conclusion and Future Work ......................................................................... 57
8 References .................................................................................................... 59
9 Appendix ....................................................................................................... 63
  9.1. IPIP Personality Questionnaire ................................................................. 63
  9.2. Probabilities and Rewards of Industries and their corresponding Groups for the Industry similarity measure ................................................................. 65
List of Figures

Fig. 1: The three defining points of a team ................................................................. 2
Fig. 2: Collection of user information ........................................................................... 3
Fig. 3: Main functions of PROTEAS .............................................................................. 19
Fig. 4: User data collection ......................................................................................... 20
Fig. 5: Team building process ..................................................................................... 21
Fig. 6: Control flow of a user’s login .......................................................................... 22
Fig. 7: Sample industries and their corresponding groups ........................................... 31
Fig. 8: Welcoming screen ......................................................................................... 39
Fig. 9: LinkedIn permissions screen ........................................................................... 40
Fig. 10: The introductory screen .................................................................................. 41
Fig. 11: Screen prompting the user to fill in the personality questionnaire ............... 41
Fig. 12: A personality questionnaire being completed ................................................. 42
Fig. 13: A user’s personality profile .......................................................................... 42
Fig. 14: A user’s profile information ........................................................................... 43
Fig. 15: A user’s connections ..................................................................................... 43
Fig. 16: Search criteria ............................................................................................... 44
Fig. 17: Personality criterion example ....................................................................... 45
Fig. 18: Results page. The first 3 results are displayed, out of 84 total ...................... 46
Fig. 19: Breakdown of general similarity (all measures selected)............................... 46
Fig. 20: Breakdown of general similarity (4 out of 6 measures selected) .................... 47
Fig. 21: Sample results after re-ranking based on education ..................................... 47
List of Tables

Table 1: LinkedIn API Permissions ................................................................. 23
Table 2: Degree weights .................................................................................. 34
Table 3: Weights of individual similarity measures ......................................... 35
Table 4: Proposed teams for Scenario #1 ....................................................... 51
Table 5: Proposed team for Scenario #2 ........................................................ 53
Table 6: Proposed teams for Scenario #3 ....................................................... 54
Table 7: Industry Groups .............................................................................. 66
Introduction

As the years go by, our way of life is admittedly getting more hectic. We have to work more, achieve more, experience more and consume more. This high consumer demand and, at the same time, the harsh international competition is pushing businesses to create new products and services, ever faster, while maintaining or, better yet, improving quality. In order to cope with this new volume of demand that defines our era, they have to utilize every possible “weapon” in their arsenal. One way to do so is to reassess one of the first and most important steps in new product development: Team Building.

Teams are more often than not created based on the technical requirements of a given task, project or product. As a result, team members are selected because they have the necessary skills, experience and expertise. Even seniority and hierarchy are sometimes taken into consideration. Research has indicated, however, that personality can play a significant role, not only to individual characterization, but also in relation to group dynamics. A team may be successful as a whole, due to the competency of its members for the task at hand, but it is suggested that the cohesion of the group and the right match of its members’ personalities, also contribute positively. The interpersonal relationships among team members may also affect team performance.

During the last decade, we have witnessed a great increase in the use of social media. People of any age and all walks of life, from all over the world, are using at least one of the many social media available, on a regular, if not daily, basis: Internet forums, blogs, wikis, social networks. Not only do we encounter a
plethora of social media, but new types also emerge, such as social networks with a business direction.

In our research, we have made the following remarks, which instigated our motivation for this dissertation:

- While new advances are noticeable in many areas, team building processes seem to remain stagnant.
- New factors have been identified, which may boost team productivity and efficiency.
- Team cohesion and good interpersonal relations contribute, not only to the team’s efficiency, but also to each individual’s well-being.
- Nowadays, social media have become very popular, and play quite a great role in our everyday lives.

We would like, for these reasons, to approach the subject of team building under a different light, and provide a way for all these new criteria to be included in the team formation process, so as to create more efficient and productive teams with greater cohesion. This will be the result of the combination of team members’ business profile (i.e. their work-related traits and skills), their social connectivity and also their personality traits (Fig. 1).

Fig. 1: The three defining points of a team
As a result of our work, PROTEAS has been created, an online application, which combines users' professional profiles acquired through LinkedIn, a business-oriented social network, their connectivity in this network, as well as their personality characteristics, which are acquired through a personality test we provide. The personality characteristics are based on the Big Five Factor Model, which regards personality as a combination of five traits: Openness to Experience, Conscientiousness, Extraversion, Agreeableness and Neuroticism. A rough outline of this procedure is presented in Fig. 2. The aim is to propose teams, which are expected to fare better than any “conventional” one. For this purpose, we have created different similarity measures, so as to compare users and rank teams, according to the specific requirements that have been set.

1.1. Open Problems

As with any research, we have encountered a couple of problems, for which we have no power or control over. These are:

- **Literature on personality**
  Although a great deal of research has been carried out, personality isn't something that one can quantify and attach a specific and
accurate numerical value to. So, it is even more difficult to talk about team dynamics and the role of personality in them. There is, of course, a decent amount of research available in that direction, which helped us make our first steps in this endeavour, but unfortunately, the hypotheses are significantly more than the solid results.

- **User connectivity**
  Social interactions and interpersonal relations are more intricate than they appear to be and are everything but black and white. Unfortunately, we don’t have a way to calculate them in their full extent and we rely on the information we can infer from the social network that we have chosen to work with. (See section 3.3 for more details).

- **Data quality**
  A final concern is the quality of the data available. As with every social network, each user is free to add any information he deems necessary. For our purpose, we would like sufficient and correct information to be provided, in order for our similarity measures (see section 4.5) to work efficiently. Of course, this is not always the case.

### 1.2. Dissertation Structure

The remainder of this dissertation is structured as follows: In Section 2 we present previous work, that deals with personality and work effectiveness. A user’s personality traits may indicate his working performance. Also, the personality composition of the whole team can accordingly indicate the team’s expected performance and efficiency. Moreover, a few other factors may influence team cohesion, such as the interpersonal relations of the team members and their social connections.

In Section 3 we present some fundamental ideas, which form the basis of our work. As we have stated, the three defining points of a team are the users’ business profiles, their social connectivity and their personality profiles. We
explain these three concepts, and we provide some general info on team and team forming.

The theory behind our method for team forming is described in Section 4. First we present the user data collection process. Through the users’ LinkedIn accounts, and after having their explicit permission granted to us, we acquire info about their professional profiles and their connectivity to other users. Through a personality questionnaire we provide, we acquire each user’s personality profile, according to the Big Five Factor Model. Then, we explain the method by which teams are created and then ranked, which leads to the suggestion of the best team available.

In Section 5 we describe PROTEAS (PROfessional TEAms in Social networks), an application which puts into action our method. A number of search criteria may be set, possible teams are created and the team with the greatest similarity among its members is proposed.

Section 6 presents the scenarios that test and validate our application. We also discuss its efficiency and usefulness and also some of the problems we have encountered.

Lastly, in Section 7 we conclude and discuss some possible future directions.
2

Literature Review

Traditional team building criteria are those concerning a team’s demographics or members’ skill-set. Team building processes of this kind are trivial and widely known. We would like to focus more on the role of personality on work efficiency and explore other factors that benefit a team’s effectiveness.

2.1. Personality and Individual Efficiency

The study of Personality started in the 50’s and the definition of the Five-Factor Model soon followed. Since then, research moved forward and focused, not only in what traits and factors constitute an individual’s personality, but also in how these traits affect him in his working environment. A great portion of research has been made in order to establish a connection between personality traits and job efficiency.

Conscientiousness has been regularly found to be a valid predictor of job performance, either in general (Hurtz & Donovan, 2000) or in specific job families: Managers, Professionals, Police, Sales People and Skilled/Semi-skilled (Barrick & Mount, 1991). In cases where job performance has been divided into subcategories in order to be studied further, it has again predicted each one equally well, such as: job proficiency, training proficiency and personnel data (Barrick & Mount, 1991), or task performance (Rothmann & Coetzer, 2003; Hurtz & Donovan, 2000), job dedication, and interpersonal facilitation (Hurtz & Donovan, 2000). As Conscientiousness describes people who are hard-working, organized and reliable, it is not unexpected that they appear to
perform better than those who believe that they lack those characteristics (Hurtz & Donovan, 2000). Conscientiousness is also, to a degree, positively associated with creativity (Rothmann & Coetzer, 2003).

Neuroticism has showed, in general, relatively low correlations, which can be insignificant (Barrick & Mount, 1991), or borderline significant, as it appears that it may have a small, but consistent, impact on job performance (e.g. in Sales and perhaps Management) and in the aspect of interpersonal facilitation (Hurtz & Donovan, 2000). A practically significant correlation has also been found with management performance. People with higher Neuroticism values tend be less creative and not perform as well as those who are emotionally stable (Rothmann & Coetzer, 2003).

Openness to Experience has been found to validly predict performance in Management (Rothmann & Coetzer, 2003) and jobs involving Customer Service (Hurtz & Donovan, 2000). It also predicts Training Proficiency relatively well (Barrick & Mount, 1991) and Task Performance (Rothmann & Coetzer, 2003). It is also related to creative and imaginative tasks (Rothmann & Coetzer, 2003), but not so, when the task at hand has a routine nature (Reilly, et al., 2002).

Extraversion is a valid predictor of job performance for Managers and Sales people (Barrick & Mount, 1991; Tett & Burnett, 2003; Hurtz & Donovan, 2000) and for those occupations that have social aspects (Kichuk & Wiesner, 1997). It also predicts, quite significantly, training proficiency (Barrick & Mount, 1991) and task performance (Rothmann & Coetzer, 2003).

Agreeableness has not been correlated with job performance (Barrick & Mount, 1991; Tett & Burnett, 2003), but further studies did find a small correlation with management performance (Rothmann & Coetzer, 2003) and interpersonal facilitation. Agreeableness could be important for jobs that require interpersonal interactions, so that being likeable, cooperative and good-natured has a small but consistent impact on performance (Hurtz & Donovan, 2000).

### 2.2. Personality and Team Efficiency

How personality traits influence the efficiency of a whole team is still a matter under discussion. Quite a few hypotheses have been proposed, but even fewer researches have been carried out either to confirm or to contest them.
As stated above, Conscientiousness has been found to be the most valid predictor for individual performance. On a team setting, along with Agreeableness, they seem to be highly significant and to take on a greater role, as working in a team involves communication and social interaction (LePine, et al., 2011; Neuman & Wright, 1999). Social interaction could also be perceived as a trait of Extraversion. Thus, Extraversion is also linked with team effectiveness (Kichuk & Wiesner, 1997). Nevertheless, there are cases, such as in innovative product design teams, where having high Agreeableness levels could be detrimental; an individual challenging the ideas of the team could lead to reassessing the problem and producing better results (Thoms, et al., 1996; Humphrey, et al., 2007). Emotional Stability, or alternatively the luck of Neuroticism, is believed to be positively correlated with team performance (Mann, 1959; Thoms, et al., 1996). Unfortunately, the role of Openness to Experience is not well defined in a team setting (Kichuk & Wiesner, 1997).

Apart from what traits are necessary to form a successful team, another aspect in research is the variance between members’ personality traits. On the one hand, homogeneity is supported, as members of a homogeneous team seem to get along better, since they share the same personality variables, while, on the contrary occasion, divisive tensions seem to rise and productivity to decline, due to a clash in personality (Tziner, 1985; Halfhill, et al., 2005). On the other hand, homogeneity is proposed for ordinary and routine tasks. For less structured tasks, a heterogeneous team will outperform any other (Basadur & Head, 2001; Reilly, et al., 2002). It should be noted however, that members of overly heterogeneous teams tend to be more dissatisfied. The same applies, to a lesser extent, to members of overly homogeneous teams (Basadur & Head, 2001).

2.3. Other Factors Relating to Team Efficiency

On another note, demographics (e.g. age, gender, seniority) seem to have no relation whatsoever, for the successfulness of a team. For instance, Kichuk & Wiesner (1997) found no significant difference between teams, with respect to the number of female members participating. Also, Reagans, et al. (2004) compared the efficiency between teams based on demographics and teams based on the social connections of its members. Demographics appear to be the easy option to create and manage teams, but ultimately there is no sign that the teams are more productive. On the contrary, the social connectivity of team
members seems to boost productivity. Kamel, et al. (2011) also note, that in a team setting, there is a communication cost between members. As teams get larger, the communication cost increases. This cannot always be avoided, as team size seems to be proportional to the number and size of the required tasks. A small or medium sized team, however, is the optimal one, in terms of communication cost.

2.4. Literature Discussion

In the last decades, focus has switched from the “traditional” team building techniques, that is, those that just take into account the task-related skills a team should have and maybe the demographics of the team, to newer techniques, which promote the cohesion and smooth operation of the team. This can be achieved through personality-based matching of team members, or by taking into consideration their social connectivity and interpersonal relations. There is no real consensus, however, about which of these methods is the most efficient, or what the most appropriate personality characteristics are, that a team’s members should have.

For this reason, we have created a framework, and provide a platform, so as to create teams, which combine both the social connectivity of their members and their personality composition. Teams can also be optimized according to one of these two characteristics. We aim to create coherent teams and see what the connection is between team composition and team efficiency.
3

Basic Concepts

In this section, a brief overview of some fundamental concepts is provided. These concepts are used throughout the text and are the building blocks of our application: A user’s professional profile, his personality and its characteristics, a social network, as well as teams and team forming.

3.1. Professional Profile

A professional profile is a collection of information mainly regarding a person’s working experience and also other qualifications. The qualifications may include, but are not limited to, the skills a person has acquired which are useful for his line of work, the foreign languages he may speak or his educational and academic backgrounds. A professional profile may also include a person’s previous teaching or research experience, publications, presentations, awards, honours, or other achievements.

In its written form, usually called curriculum vitae (CV) or résumé, it is almost exclusively used in order for people to apply for a job position. Employers can thus quickly interpret the experience and qualifications of each applicant. In most cases, the professional profile is used as a first step, a pre-selection process, so that a formal interview may follow. As such, it has to elicit the desire for the employer to meet the candidate in person. For this purpose, many theories and useful tips exist, to aid potential candidates in
creating compelling professional profiles, but this subject is well beyond the scope of this dissertation.

In the Information Age that we live in, professional profiles have gone through a transformation, and are now digitally created and distributed. As a direct result, many websites have been created to help people with this process. Some of them take the form of fully fledged social networks, which provide quite a few tools for their users. Users of these networks are able not only to create their professional profiles online, but also to easily edit them and keep them up-to-date, and even apply on the spot for various positions. We explain more about social networks, and LinkedIn in particular, a social network with a professional orientation, in Section 3.3.

In the context of this dissertation, we use professional profile as a term describing a person’s collective skills, knowledge and other work-related competencies.

3.2. User Personality

The simplest definition of personality can be found in a dictionary: “The combination of characteristics or qualities that form an individual’s distinctive character” (Simpson & Weiner, 1989). In the field of Psychology, from time to time, many different theories have been proposed, about what is personality, what are its parts, how it is affected, and how it affects us (e.g. by Sigmund Freud, Raymond Cattell, Hans Eysenck, Heinz Hartmann, Albert Bandura, etc). Each one is strongly influenced by its author’s ideological and methodological background. However, even if these theories have their strong points and manage to get ever closer to the idea of personality, in recent years they tend to be overcome by a new theory, which is agreed upon by most of the modern personality psychologists. This is the Big Five Factor Model (John & Srivastava, 1999), whereby each person’s personality can be described as a combination of five major traits, or dimensions. These are the Big Five Personality Dimensions: Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism (Easily remembered through the “O.C.E.A.N.” acronym). Each dimension is described below, and the way they are measured and calculated follows.
a. **Openness to Experience** or **Intellect** is a person’s curiosity, imagination and creativity. People with high Openness levels tend to find routine work tiring and prefer a variety of tasks. They are also creative, imaginative, curious, with unconditional ideas, in touch with their feelings and greater artistic sensibility. People with low levels of Openness tend to be more realistic, with traditional ideas, a bit conventional, preferring simple and obvious situations with strong resistance to change.

b. **Conscientiousness** is linked with organization skills, work ethic, self-discipline and goal realization. Conscientious people achieve their goals with persistence and thorough planning and their actions are not usually driven by spontaneity and impulsiveness. Those with low conscientiousness levels tend to be disorganized, undependable and negligent.

c. **Extraversion** is the level of sociability and enthusiasm, and is in a way a measure of the intensity of the person’s interaction with others. People with high Extraversion levels seek the company of others, are enthusiastic and full of energy, driven by the urge for action, thrill and excitement. They are sociable and a bit of attention seekers. On the contrary, people with low Extraversion levels are more introverted and tend to be quite, withdrawn and more detached from the social life, as they need time for themselves, without the need for the company of others.

d. **Agreeableness** indicates how kind, altruistic and friendly is one person towards others. People with high Agreeableness levels tend to be cooperative, compassionate, forgiving, non-competitive, generous and willing to help others or to step down in case there is a clash of interests. They are mostly sanguine and believe in the best of each person. For people with low Agreeableness levels personal interest takes precedence over their social behaviour, as they don’t need to get along with others. They are uncooperative, critical, rude, harsh and can become very competitive.

e. **Neuroticism** or **Emotional Stability** indicates whether a person is tranquil and calm or irritable, emotionally unstable or moody. People with high Neuroticism tend to have negative feelings, such as anger, rage, stress, or depression. They can be pessimistic, insecure, gloomy and exaggerate when they come up against a problem or difficult
situation. On the contrary, people with low Neuroticism are calm, emotionally stable, relaxed, lacking stress and anxiety.

A person’s personality profile is, more often than not, assessed through a personality test. A personality test usually contains a number of questions (de Raad & Perugini, 2002), where each question corresponds to a specific personality dimension. The number of questions could range from 10 (Gosling, et al., 2003) to 300 (Johnson, 2000), or even more (Schmit, et al., 2002). Questions present to the user an attitude towards a subject, a situation, or a feeling, such as “I appreciate art”, “I feel comfortable around people”, or “I seldom feel blue”. The answers follow the form of a Likert-type scale. In a Likert-type scale, for each question respondents specify the degree of their agreement or disagreement on a symmetric agree/disagree scale. For example if there are six available options to answer from, three of them are levels of disagreement, while the other three are levels of agreement (e.g. “I completely agree/disagree”, “I agree/disagree”, “I barely agree/disagree”). A neutral option might or might not be available in the middle (“I neither agree, nor disagree”).

Each answer corresponds to a specific value, which depends on the number of possible answers and the positive or negative correlation of the question. For example, if the test uses 4 possible options (Strongly disagree, disagree, agree, strongly agree), each option has a value of 1 to 4. If the question has a positive correlation to the specific personality dimension (e.g. “I make friends easily”, for Extraversion) the values are:

- strongly disagree: 1
- disagree: 2
- agree: 3
- strongly agree: 4

On the contrary occasion (e.g. “I don’t make friends easily”) the values are the opposite, i.e.:

- strongly disagree: 4
- disagree: 3
- agree: 2
- strongly agree: 1
These values are then added separately for each personality dimension and then normalized to the desired scale.

We are using a test with 50 questions, which can be answered with 1 out of 5 available answers, as described in detail in Section 4.3.

### 3.3. Professional Social Networking: The Case of LinkedIn

According to boyd & Ellison (2007), a social network could be defined as:

“A web-based service that allows individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system”.

Social networks have started appearing in their current form about a decade ago (Skeels & Grudin, 2009). Since then, their popularity and use has increased exponentially. Each social network usually has some or all of the following characteristics (Mayfield, 2008):

- **Participation.** Every user and member of a social network can participate by making contributions or giving feedback to anyone who is interested.

- **Openness.** Access to information is open. Users are encouraged to share information and experiences and comment on other people’s shared content.

- **Conversation.** While the norm has been to “broadcast” information, meaning that content is distributed to an audience, social networks employ a two-way conversation between users.

- **Community.** A sense of community permeates social networks, as users who share common interests, ideas or hobbies may congregate and communicate effectively.

- **Connectedness.** The majority of social networks is based on user connectivity. Users connect to each other and they also link to other sites, content and resources.
A social network may have a specific context and, for example, may bring together people with common music interests (e.g. MySpace), people who are searching for friends (e.g. Facebook) or people who want to promote their business profile (Ellison, et al., 2007). There are quite a few business-oriented social networks, such as LinkedIn, Upspring, Xing, and PartnerUp. LinkedIn is the most recognizable network of its kind, with 225 million members from 200 countries².

LinkedIn focuses on professional information and users are encouraged to create their professional profile online. They can, thus, showcase their working experience, the skills they have acquired that are essential for their professional field, their studies and diplomas, the foreign languages they have studied, as well as other academic experiences and activities they have taken part in. The image that LinkedIn conveys is strictly professional. There is no way for a user to add information for example about hobbies, political or religious beliefs, or types of entertainment that he may like. This is also reflected on a user’s profile page. The content of the page is mostly static, looking almost like a CV, and the highlights are the user’s current working position and the industry he belongs to.

The second major aspect of LinkedIn is a user’s in-network contacts, called “connections”. These are people they know and trust within their line of work with whom they can communicate, trade information and mutually refer. The referral is a system where users “vouch” for a connection’s competence and endorse any of his posted skills. LinkedIn employs a “gated-access approach” (Papacharissi, 2009), which means that users have access only to their direct connections’ list of connections. If they want to connect with any other, a mutual connection has to act as an intermediary. This is done in order to inspire confidence and trust among users.

Access to the site is free for anyone, although professionals or graduating students tend to sign up most (Papacharissi, 2009). After all, LinkedIn states on its front page that it is a “network of professionals”. LinkedIn users can be divided into 3 categories (Skeels & Grudin, 2009): current or recent students, young professionals and older professionals. LinkedIn is not targeted to younger students, as it is in a way not useful to them. As we have already pointed out, the major feature of a user’s profile is his working experience and current job title, which many students lack. Older professionals who already have established careers don’t find LinkedIn too enticing, but a form of peer

² http://www.linkedin.com/about-us, retrieved August 2013
pressure forces them to create accounts in order to connect with other users. Young professionals find LinkedIn more useful, as their profiles act as a concise and updated CV, always available to many prospective employers. The same, albeit to a lesser extent, applies to recent students.

LinkedIn also provides an Application Programming Interface (API), in order for developers to create applications. The popularity of the site, its business orientation, and the ability to gather user data through the updated API, were the main reasons that LinkedIn was our first choice in helping us implement our application.

3.4. Team and Team Forming

A team is a group of people working interdependently, with a full set of complementary skills, required to successfully carry through a given task, job, or project, for which they are mutually accountable (Katzenbach & Smith, 1992). According to the BusinessDictionary (WebFinance, Inc., 2013), team members:

“(1) operate with a high degree of interdependence, (2) share authority and responsibility for self-management, (3) are accountable for the collective performance, and (4) work toward a common goal and shared rewards. A team becomes more than just a collection of people when a strong sense of mutual commitment creates synergy, thus generating performance greater than the sum of the performance of its individual members”.

A team should not be confused with a working group. A team may be, in essence, a working group, but a working group is not a team. A working group is just a collection of individuals, where each participant has his own goals and takes individual responsibility for the task that he alone has undertaken. There is no sense of real cohesion and member communication is more formal (Katzenbach & Smith, 1992). Within a team on the other hand, members need to cooperate in order to accomplish their tasks, and they are responsible for a whole project or product, rather than its parts (Verzuh, 2011).

The composition of a team is usually suggested by the size or the difficulty of the task at hand, the time allotted for it, and the people currently available.

3 http://developer.linkedin.com/
In most cases, team members are either chosen because they have the necessary skills or work experience (usually in work environments; a manager may be responsible for creating the team, and he chooses individuals who are known to perform well and be efficient) or because the members have some degree of previous acquaintance and there is a strong interpersonal fit between members (as observed in academic environments; students who are given a group task or a group assignment choose teammates among their friends first, regardless of each one’s experience or knowledge) (Aldrich & Kim, 2007). In total, four stages of team development can be observed (Tuckman, 1965):

- **Forming.** The team is newly formed. Members get acquainted and there is a relative calmness and harmony between them.

- **Storming.** As team assignments are being negotiated and discussed, conflicts and disagreements begin to arise concerning the optimal course of action.

- **Norming.** After a while, rules of conduct between members have been established and they learn how to work together in harmony. This is the moment when the sense of a real team appears.

- **Performing.** The team works effectively, without problems and conflicts. This stage can amount to as little as 25% of the total team life cycle.

In order for a team to be efficient, team members should have complementary skills, specific goals to strive for, which allow mutual accountability and be composed of a small number of people (e.g. 5-7) so that consensus without discord is ensured (Gordon, 2002).

Successful teams should have members who are fully committed to the team project or task, trust each other, communicate openly and freely, providing fresh ideas, are able to assess their work continuously and are not afraid to take risks. It is also advisable to have different types of team members (e.g. a leader, a hard worker, a good thinker, etc.) so that a team can work efficiently (for example, it isn’t advisable for a team to have more than one leader) (Verzuh, 2011).
PROTEAS: “Automated generation of PROfessional TEAms in Social networks”

What we are trying to do is to propose a method, which will automatically create teams. Teams are usually formed by a supervisor, who has a specific task or a final goal in mind and team members are selected accordingly. Our objective is to introduce non-traditional variables, such as the social connectivity of the team members and their personality information, in order to create a team that will perform better and more efficiently. To this end, we have created PROTEAS, a team building application.

4.1. PROTEAS Methodology and Principles

As the name suggests, PROTEAS is an application designed to automatically create teams. Its purpose is essentially twofold: To collect user data and present them to the user and to propose teams based on these data (Fig. 3).
To connect to our application, a user must have a LinkedIn account, so that he can use his credentials and log in to PROTEAS. We then proceed to draw his professional profile information with the help of the LinkedIn API, after of course acquiring his permission to do so. This process is explained in more detail in Section 4.2. We also provide a personality questionnaire, which the user is able to complete, after successfully logging in. The questionnaire parameters are explained in Section 4.3. The user data collection process is presented in Fig. 4.

The second part of PROTEAS is the team building algorithm. As we have already discussed, teams are no longer formed based strictly on the skills and knowledge of the available candidates. Their interpersonal relations, as well as prospective teams’ personality composition play a major role. We want to reflect this new trend in our application and we offer the option to include these parameters during team formation. All of the available team parameters are explained in more detail in Section 4.4.

Our team building method is presented in Fig. 5. The criteria that each team should have are set beforehand, and afterwards a search is initiated. All matching users are fetched from the database. We proceed to create all possible teams, according to the specified team size. Teams are then ranked, based on a formula we have devised, which computes the similarity between the users of each team. This value is representative of the whole team, and they can
therefore be ranked accordingly. The similarity formula, and the way that it is created, is presented in Section 4.5 and in Section 4.6 we present our ranking algorithm.

4.2. Application Front-End and User Data Collection

In order for someone to use our application, he must be a LinkedIn member. Using his LinkedIn credentials he can log in to Proteas. The process of logging in or signing up is presented in Fig. 6. In accordance with LinkedIn policy, the user has to grant permission for our application to use his LinkedIn information.

The LinkedIn API provides 9 different types of permission. Any of them, or even all of them, can be chosen if needed by the application being developed. We have deemed only 3 permissions necessary for our application, which give us access to the following information:

- Basic Profile Overview, including name, photo, industry and current working position, if applicable.
- Full Profile Information, including past working experience, education, skills, spoken languages, and current location.
- Connections, providing information about the user’s 1st and 2nd degree connections.

All nine permission types are presented in Table 1.
After permission is granted, a call is made to the LinkedIn API and the user's information is copied to a MySQL database. This action is done only once, during the user's first visit. Consecutive visits require just a login.

The application has a pretty simple interface. Upon logging in, the user is presented with his business profile. This information is fetched from the database and corresponds to the data drawn from LinkedIn.

The user has the option to fill in a questionnaire, in order to assess his personality outline, according to the Big Five Dimensions. The questionnaire is described in the following section.

The major feature of our application is the team building tool, which is described in detail starting in Section 4.4.

4.3. Personality Questionnaire

Ever since Personality has started to be explained as part of human psychology, various tests have been created, such as the Woolworth
### Table 1: LinkedIn API Permissions

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Permission</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>r_basicprofile</td>
<td>Profile Overview</td>
<td>Name, photo, current positions</td>
</tr>
<tr>
<td>r_fullprofile</td>
<td>Full Profile</td>
<td>Past working positions, education, skills, languages, location</td>
</tr>
<tr>
<td>r_emailaddress</td>
<td>Email Address</td>
<td>Primary email address</td>
</tr>
<tr>
<td>r_network</td>
<td>Connections</td>
<td>1st and 2nd degree connections</td>
</tr>
<tr>
<td>r_contactinfo</td>
<td>Contact Info</td>
<td>Home address, telephone number</td>
</tr>
<tr>
<td>rw_nus</td>
<td>Network Updates</td>
<td>Retrieve and post updates to LinkedIn as the user</td>
</tr>
<tr>
<td>rw_company_admin</td>
<td>Company Page and Analytics</td>
<td>In case the user is an administrator in a company page, this permission offers the possibility to edit such pages and post status updates on behalf of the user</td>
</tr>
<tr>
<td>rw_groups</td>
<td>Group Discussions</td>
<td>Retrieve and post group discussions as the user</td>
</tr>
<tr>
<td>w_messages</td>
<td>Invitations and Messages</td>
<td>Send messages and invitations in order to connect with other users</td>
</tr>
</tbody>
</table>

Personality Data Sheet, developed during World War I (Nezami & Butcher, 2000), the Myers-Briggs Type Indicator, based on theories of Carl Jung (Briggs, 1976), the Minnesota Multiphasic Personality Inventory (MPPI), used by many organizations worldwide such as the C.I.A. (Hathaway & McKinley, 1951), and the Rorschach Test, which employs inkblots as a tool for personality assessment (Klopfer, 1951). Nowadays, some of the personality tests pertaining to the Big Five Factor Model are the following:

- Revised NEO (Neuroticism-Extroversion-Openness) Personality Inventory (NEO PI-R) with 240 items, one of the most significant (Costa & MacCrae, 1992)
- Five Factor Personality Inventory (FFPI) with 100 items (Hendriks, et al., 1999)
- International Personality Item Pool (IPIP) (Goldberg, 2001)
- FFPI for Children (FFPI-C) (McGhee, et al., 2007)
- Structured Interview for the Five-Factor Model of Personality (SIFFM) with 120 items (Trull, 1997)
- Big Five Inventory (BFI) with 60 items (John & Srivastava, 1999)
- Ten-Item Personality Inventory (TIPI) with just 10 items (Gosling, et al., 2003)

Most of these tests are used commercially, and are provided after a fee has been paid. As such, we selected IPIP – the International Personality Item Pool⁴, which is a public domain list of items and thus provided gratis, covering the Big Five personality traits. Moreover, as the name suggests, there is a pool of available items, and the total number of items a test has, can be adjusted according to anyone’s needs.

An extended personality test can produce more fine-grained and accurate results (Nunes, 2008). The tradeoff is that, due to its length, it takes quite some time to complete, a fact that discourages users. In the case that too few questions are presented, the problem lies in the fact that results are too coarse-grained and are, therefore, not particularly useful. We set the number of items to 50, as a compromise between test length and user willingness to spend time in order to complete it. These 50 items are presented in Appendix 9.1. The test is a typical 5-point Likert-type scale (as already explained in Section 3.2), where each item can be answered with one of the following options:

1. Strongly disagree
2. Disagree
3. Neither disagree nor agree
4. Agree
5. Strongly Agree

⁴ http://ipip.ori.org/
Every item presents the user with a situation or belief. There are 10 items for each of the 5 personality dimensions.

4.4. Team Building: Search Criteria

In order for a team to be created, some information must be known beforehand, which is provided by setting the appropriate search criteria. The search criteria that we provide correspond to the information we acquire through each user’s LinkedIn account and their personality questionnaire. The most basic is the number of members the team should have (Defined as r). The remaining available search criteria are the following:

1. The **skills** each member of the team should have.
2. The **languages** they should speak.
3. Their **location**, defined as specific as a city (where LinkedIn provides the appropriate information), or as broad as a continent.
4. The minimum –or maximum– required level of **academic background**.
5. One or more **industries**, in which team members may belong to.
6. The **social connection** between them, in accordance with their LinkedIn profile connectivity.
7. The **personality** that the members of the team should have. The personality is defined through the Big Five Dimensions (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism). For each dimension, a relative value can be set (low, medium, high). Moreover, each dimension can be disabled individually, in case a user wants to make a more specific search, or if he considers that a specific dimension is insignificant or irrelevant at the moment.

A search in the database takes place, selecting users who match all the criteria that have been set, except for the social connectivity one, which is used later on in the process; social connectivity is a measure between two or more users and cannot apply to a single user directly. It is presented and
explained in more detail in Section 4.6. If the users fetched from the database (n) are less than or exactly the number that was set as the team size (n ≤ r), all matching users are presented as a possible team. On the contrary occasion, where more users are fetched (n > r), all possible teams are calculated: By having n total available users, and trying to create a team with r members, all possible combinations are:

\[ \text{Combinations}(n, r) = \frac{n!}{(n - r)! \cdot r!} \]

Then a process of ranking is employed, so as to propose the best team available. For this purpose, a metric must be defined, in order to compare the available teams. In the following section we describe the steps that lead to this metric.

### 4.5. Similarity Measures

In order to be able to rank teams, each one must be represented by a value. For our purpose, this value will denote the similarity among members of a team. We start by calculating the similarity between two members of the team, and in the following section we describe how the whole team’s similarity is computed and how the teams are ranked.

In order to calculate the similarity between two different users, we provide a set of different similarity measures. By default, all of these measures are taken into account while computing the general similarity between two team members, but the user has the option to disable any of them. The available similarity measures correspond to the first six search criteria: Skills, languages, and industries, which are nominal values, and location, education, and social connectivity, which are ordinal values. Nominal values are those who may belong to one or more sets and have a descriptive property of the data (e.g. male, female). Ordinal values are distinct and denote the ability to establish a ranking over a set of values (e.g. distance, temperature, low-medium-high values, etc).

For each category containing ordinal data, we have devised a specific formula, which is described below. For nominal data, we must make set operations. The most common similarity measure for sets is the Jaccard Index (or Jaccard Similarity Coefficient), which is defined as follows:
Consider two sets, $A$ and $B$, where:

\[
A = \{a_1, a_2, \ldots, a_{n-1}, a_n\}, |A| = n
\]

\[
B = \{b_1, b_2, \ldots, b_{m-1}, b_m\}, |B| = m
\]

Then:

\[
\text{Jaccard Index}(A,B) = \frac{|A \cap B|}{|A \cup B|}
\]

The basis for our formulae is the Overlap Coefficient:

\[
\text{Overlap Coefficient}(A,B) = \frac{|A \cap B|}{\min(|A|, |B|)}
\]

which is based on the Jaccard Index and produces a bit bigger and, for our purpose, better results.

Before proceeding with the description of each similarity measure, we will create two “test subjects”, user A, and user B, each with his own characteristics, in order to showcase each similarity measure based on these two users.

- **User A**
  - **Skills**: Algorithms, Accounting, Databases, Programming.
  - **Languages spoken, with level of proficiency, where available**: English (native or bilingual), Greek (full professional), Japanese (limited working), Swahili.
  - **Industries he belongs to**: Main: Biotechnology. Secondary: Internet.
  - **Geographical location**: Greece.
  - **Degrees acquired**: Master’s, Doctorate.
User B

- **Skills**: Algorithms, Machine Learning, Programming

- **Languages spoken, with level of proficiency, where available**: English (limited working), Greek (full professional), Japanese, Swahili, Korean.

- **Industries he belongs to**: Main: Civil Engineering. Secondary: Telecommunications.

- **Geographical location**: Japan

- **Degrees acquired**: Bachelor’s, Master’s.

### 4.5.1. Skill-based Similarity Measure

Each user can define a set of skills that he has acquired. To find the similarity between two users’ skills, the overlap coefficient is used:

$$Similarity_{skill}(A, B) = \frac{|Skills(A) \cap Skills(B)|}{\min(|Skills(A)|, |Skills(B)|)}$$

From our example, we have:

- $Skills(A) = \{\text{Algorithms, Accounting, Databases, Programming}\}$
- $Skills(B) = \{\text{Algorithms, Machine Learning, Programming}\}$

Thus:

$$Similarity_{skill}(A, B) = \frac{|\{\text{Algorithms, Databases}\}|}{|Skills(B)|} = \frac{2}{3} = 0.66$$

### 4.5.2. Language-based Similarity Measure

In defining a spoken language, the user can optionally set the level of his proficiency. There are five available levels: Native or Bilingual, Full Professional, Professional Working, Limited Working, and Elementary. Since not all users provide this information, we consider a language rated, if both users have set a level. If only one user has rated a language, the level is
discarded. To make use of the extra information available, we use an altered version of the Overlap Coefficient.

By calculating the intersection of two sets, each common element counts as 1. So, common languages of two users, which are not rated, should contribute with a value of 1.

For the rated languages we use a more fine-grained approach, where the common language does not count as 1, but can have a value between 0 and 1, according to how close the levels of each user are. In order to do so, we start by converting the proficiency levels to numbers. The lowest proficiency level has a value of 1, and the highest a value of 5. In this way, we have a numeric representation of a user's language proficiency level. Next, we define the language proficiency range as:

\[ \text{range} = \text{maxRate} - \text{minRate} + 1 = 5 \]

The value between two rated languages can be calculated as:

\[
\text{value}(\text{Language}(A), \text{Language}(B)) = \frac{\text{range} - \text{diff}}{\text{range}} = \frac{5 - \text{diff}}{5}
\]

where

\[ \text{diff} = |\text{rate}(\text{Language}(A)) - \text{rate}(\text{Language}(B))| \]

So, if two users speak the same language with the maximum difference in proficiency levels (1 and 5 respectively), the value for that particular language is calculated as:

\[
\frac{5 - 4}{5} = \frac{1}{5} = 0.20
\]

while, in case the level is the same, the difference is 0, and thus the value is calculated as 1.

The final similarity formula is:

\[
\text{Sim}_{\text{language}}(A, B) = \frac{\sum_{i,j} \left( \text{value} \left( \text{Language}_i(A), \text{Language}_j(B) \right) \right)}{\min(|\text{Languages}(A)|, |\text{Languages}(B)|)},
\]
where

\[ \text{Language}_i(A) = \text{Language}_j(B), \]
\[ 0 < i \leq |\text{Languages}(A)|, \]
\[ 0 < j \leq |\text{Languages}(B)| \]

Considering our example (a zero value denotes no level):

\[ \text{Languages}(A) = \{\text{English}: 5, \text{Greek}: 4, \text{Japanese}: 2, \text{Swahili}: 0\} \]
\[ \text{Languages}(B) = \{\text{English}: 2, \text{Greek}: 4, \text{Japanese}: 0, \text{Swahili}: 0, \text{Korean}: 0\} \]

\[ \text{Similarity}_{\text{language}}(A, B) = \]
\[ = \frac{\text{value(English)} + \text{value(Greek)} + \text{value(Japanese)} + \text{value(Swahili)}}{|\text{Languages}(A)|} \]
\[ = \frac{5 - 3}{5} + \frac{5 - 0}{5} + 1 + 1 = \frac{3.4}{4} = 0.85 \]

4.5.3. Industry-based Similarity Measure

LinkedIn defines 147 different industries, which can belong to one, two, or three of 17 general groups. Each user may fill in an industry that he believes describes him best. We define this industry as the main industry. Additionally, a user is able to fill in his current or past job positions. Each position corresponds to a company and each company can be defined by industries. We define these as secondary industries.

Comparing two industries we can find them being the same, where their similarity is obviously 1, belonging to the same general group or groups, belonging to some of their respective groups, or belonging to different groups entirely. For example, “Biotechnology” is an industry belonging to 3 different groups: Health, Technology and Government. “Defense & Space” and “Telecommunications” belong to 2: Government and Technology. “Civil Engineering” also belongs to 2: Government and Construction. “Internet” belongs to 1: Technology. We present these sample industries in Fig. 7.
For each pair, we calculate the Overlap Coefficient of their corresponding groups. For example:

\[
\text{overlap coeff. (Biotechnology, Civil Engineering)} = \frac{|\{\text{Government}\}|}{|\{\text{Government, Construction}\}|} = \frac{1}{2} = 0.50
\]

We also define the probability for each of these cases. For example, the probability that “Defense & Space” and “Telecommunications” correspond to two users respectively, and thus both belong to the general groups Government and Technology, is 1/136 (2 same groups, out of 17 possible), while the probability between “Telecommunications” and “Internet” (having 1 common general group, out of 17 possible) is 2/17. For any such event, there is a corresponding reward:

\[
\text{reward} = 1 - \text{probability(event)}
\]

For the above examples, the rewards would be 135/136 and 15/17 respectively. In Appendix 9.2 the probability of each event, along with the overlap coefficient and the final score, is presented.

So, the score between two industries is:

\[
\text{score(Industry}(A), \text{Industry}(B)) =
\begin{cases} 
1, & \text{Industry}(A) = \text{Industry}(B) \\
\text{overlap coefficient} \cdot \text{reward}, & \text{Industry}(A) \neq \text{Industry}(B)
\end{cases}
\]
Since the main industry is a self-description of the user, we believe that it is quite important and give it a weight of 0.35. Each secondary industry has a weight of 0.15. The total score is:

\[
TotalScore_{industry}(A,B) = 0.35 \cdot score_{mainIndustry}(A,B) + \\
+ k \cdot 0.15 \cdot \sum_{i}^{k} score_{secondaryIndustry_{pair}}(A,B)
\]

The final score is normalized according to the sum of weights. So, the final similarity formula is:

\[
Similarity_{industry}(A,B) = \frac{TotalScore_{industry}(A,B)}{0.35 + k \cdot 0.15}
\]

Considering our example:

MainIndustry(A) = Biotechnology
MainIndustry(B) = Civil Engineering

SecondaryIndustry(A) = Internet
SecondaryIndustry(B) = Telecommunications

Thus:

\[
score_{mainIndustry}(A,B) = 1 \cdot \frac{15}{17} = 0.8824
\]

\[
score_{secondaryIndustry}(A,B) = 0.5 \cdot \frac{47}{68} = 0.3456
\]

And:

\[
Similarity_{industry}(A,B) = \frac{0.35 \cdot 0.8824 + 0.15 \cdot 0.3456}{0.35 + 1 \cdot 0.15} = 0.7214
\]

### 4.5.4. Similarity Measure based on Geographic Location

Two users can be geographically similar when they have a relative distance between them. According to the LinkedIn data available, we have extracted 4 possible distances. These are, from smallest to greatest (and thus denoting higher to lower similarity):
\[
\text{Similarity}_{\text{geographical}}(A, B) = \\
= \{\text{same city, same country, same continent, no similarity}\}
\]

As such:

\[
\text{Similarity}_{\text{geographical}}(A, B) = \\
= \text{Similarity}_{\text{geographical}}(\text{Greece, Japan}) = \text{no similarity} = 0
\]

4.5.5. Similarity Measure based on Educational Background

In order to calculate how similar two people are in relation to their academic background, we take into account the number and types of degrees that each user has provided. There are 5 available types, which are: Associate, Foundation, Bachelor’s, Master’s, and Doctorate. Since not every user provides complete information and may, for example, only provide his Master’s Degree, we make the following, most basic, assumptions:

1. If the user has only provided his Master’s Degree or Doctorate, we assume that he has at least one Bachelor’s Degree in his possession. This stems from the fact that a student, after acquiring his Bachelor’s Degree, can move on to further studies. Primarily for a Master’s Degree, but also in exceptional cases, straight for a Doctorate.

2. If the user has provided both a Master’s and a Doctorate Degree, we again assume that he has done so, after acquiring first a Bachelor’s Degree.

For every type of degree, we have assigned a specific weight, which is a combination of the average years required for it and an extra factor according to their level. The weights of each type of degree are presented in Table 2.

So, the score of a user’s education is calculated by adding the corresponding values of his degrees. For the two users of our example we have:

User A: \(\text{score(Bachelor’s + Doctorate)} = \)

\[= \text{score(Bachelor’s + Master’s + Doctorate)} = 4 + 3 + 7 = 14\]
Table 2: Degree weights

<table>
<thead>
<tr>
<th>Degree</th>
<th>Years</th>
<th>Factor</th>
<th>Final Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Associate/Founder</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Bachelor’s</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Master’s</td>
<td>2</td>
<td>1.5</td>
<td>3</td>
</tr>
<tr>
<td>Doctorate</td>
<td>4</td>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>

User B: \(\text{score(Bachelor’s + Master’s)} = 4 + 3 = 7\)

Finally, the educational similarity between two users is:

\[
\text{Similarity}_{\text{educational}}(A, B) = \frac{\min (\text{Score}(A), \text{Score}(B))}{\max (\text{Score}(A), \text{Score}(B))}
\]

And for our example:

\[
\text{Similarity}_{\text{educational}}(A, B) = \frac{7}{14} = 0.5
\]

4.5.6. Similarity Measure based on Social Connectivity

In social networks, two users may have \(n\) degrees of separation, where \(n\) is the number of contacts we must traverse, so that these two users can be connected through their contacts. Thus, if two users are connected they are 1 degree apart, while if they have a specific common contact they are 2 degrees apart. Since these are the only two degrees of separation we can calculate for all users of our database, we use these two, along with a third, which denotes that two users have no social connection at all. Thus:

\[
\text{Similarity}_{\text{Social}}(A, B) = \{1st\ \text{degree, 2nd\ degree, no\ degree}\} 
\]
4.5.7. General Similarity Formula between Two Users

The final similarity formula, comparing two users is defined as:

\[ \text{Similarity}(A, B) = \sum_i w_i \cdot \text{Similarity}_i \]

where \( \text{Similarity}_i \) is one of the aforementioned similarity measures and \( w_i \) is the weight of each measure. We have given different weights to each similarity measure, because we believe that some should have a greater role while defining a team. The values of each weight are presented in Table 3.

The user has the option to disable any of the similarity measures. In that case, the final similarity value is normalized according to the remaining weights.

In this way, the two users of our example have the following general similarity value:

\[
\begin{align*}
\text{Similarity}(A, B) &= 0.20 \cdot 0.66 + 0.15 \cdot 0.85 + 0.15 \cdot 0.7214 + 0.15 \cdot 0.5 + 0.15 \cdot 0 + 0.20 \cdot 1 = \\
&= 0.6427 = 64.27\%
\end{align*}
\]

<table>
<thead>
<tr>
<th>Similarity Measure</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill</td>
<td>0.20</td>
</tr>
<tr>
<td>Language</td>
<td>0.15</td>
</tr>
<tr>
<td>Industry</td>
<td>0.15</td>
</tr>
<tr>
<td>Geographical</td>
<td>0.15</td>
</tr>
<tr>
<td>Educational</td>
<td>0.15</td>
</tr>
<tr>
<td>Social</td>
<td>0.20</td>
</tr>
</tbody>
</table>
4.6. Team Similarity and Ranking Algorithm

As mentioned before, the social connectivity between two users can be described in degrees. A 1st degree denotes that two users are directly connected, a 2nd degree denotes that they both have a 1st degree connection with a third user, and in all other cases there is no connection. So, if the social connectivity option has been set during search, it can be applied at this point, where the teams have been formed and we can calculate social distances between users. Subsequently, for each possible team, the social connections of its members are checked in pairs. If there is a team which does not fulfill the set criterion, it is removed. For example, the user may have designated that all members in the team should be connected in the first degree (i.e. be directly connected with each other). A member not being connected directly “breaks” the rule, and thus the whole team is removed.

The next step is to calculate the similarity between all members of the team. Our first thought was to calculate this value taking into consideration all members of the team simultaneously. After all, three of the similarity measures are set operations. For example, to calculate the overlap coefficient of the skills of the whole team, all we need is the intersection of all skills and the smallest set of a team member’s skills. But, in fact, even the other two measures, language and industry similarity, are not solely computed through the overlap coefficient. The same applies for the other three measures, which are ordinal variables (geographical, educational and social similarities).

We could calculate these values for each pair within the team and then find the average of these values. The problem lies in the fact, that this is a very costly operation, considering that this process is repeated for every possible team and the pair values are calculated over and over again. The cost ends up increasing exponentially, as the number of team members also increases.

So, we decided to adjust this method to a simpler form. We start by calculating the similarity between each possible pair among the search results. Then, we form the required teams and for each team we calculate the average of the similarity values of each possible pair within the team. For example, let us consider a team, with four members: A, B, C, and D. The team similarity is calculated as the average of all pairs: AB, AC, AD, BC, BD, and CD. We also encountered this method in the work of Reagans, et al. (2004), where, in order to calculate a value denoting the social connectivity of a team, they took the average of each pair value within the team.
So, for all teams $T$ of size $n$, the ranking value of each team $t_k \in T$, $1 \leq k \leq |T|$, is calculated as:

$$Rank(t_k) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{Similarity}(m_i, m_j)}{\text{combinations}(n, 2)}$$

where

$$m_i, m_j \in t_k$$

We remind that:

$$\text{Combinations}(n, 2) = \frac{n!}{(n - 2)! \cdot 2!}$$

At last, the teams are sorted according to their similarity and presented in a descending order. The user has also the option to re-rank the teams according to a specific similarity measure.
5

PROTEAS Implementation

As already stated, our proposed team building method is embodied in PROTEAS, an online application we have created. In this section we present its visual interface. PROTEAS is, in a way, divided into two halves: The first contains the current user’s personal information taken from LinkedIn, and his personality characterization, as inferred from the personality questionnaire. The second is the team building tool, which essentially is a search engine. Both parts are described below.

5.1. Current User Interface

On the front screen the user is prompted to connect with our application through LinkedIn (Fig. 8). This is accomplished by using his LinkedIn credentials. The user’s LinkedIn user-ID (uid) is employed as the pivotal element, by which he is recognized. For the time being, there is no option to sign up and create an account from scratch. All data are drawn from LinkedIn.

Welcome to Proteas!

In order to begin you must be connected to LinkedIn.

Connect to LinkedIn

Fig. 8: Welcoming screen
Before continuing any further, and in compliance with LinkedIn policy, the user is presented with a dialogue, requesting his permission to share his profile information. The user is informed of the types of permissions that our application requests (as presented in Fig. 9, and already explained in Section 4.1). Two types of permissions are actually shown, as the permission for the user’s basic profile information is included in the more general permission for the user’s full profile information.

By clicking on the “Allow access” button, the process of data gathering is initiated. A preliminary and most basic API call is made, requesting the current user’s uid, in order for his identification to take place. If his uid is matched against an entry in our database, it means that the user has already visited the application before and his data have been fetched and saved. On the contrary occasion, that is, if the user uid is unknown and encountered for the first time, a second call is immediately made to the LinkedIn API, requesting all of his available data that are important for our application: Skills, languages,
industries, location, education, network connections. Once these data are gathered, they are stored on the database.

Following a successful login, the user is presented with the home screen, which is broken down into 4 main subsections:

1. The introductory screen, which thanks the user for his participation and prompts him to fill in the personality questionnaire (Fig. 10).

2. The personality questionnaire screen. This subsection asks the user to complete a questionnaire concerning his personality (Fig. 11). If the user agrees to do so, he is redirected to the questionnaire page.
Fig. 12: A personality questionnaire being completed

(Fig. 12). The questionnaire is a 50-question Likert scale test and it takes about 7 to 10 minutes to complete. The 50 questions are presented in Appendix 9.1. Upon completing the questionnaire, the user is redirected to the home screen. Now, the personality subsection presents the results of the test in a bar chart, where each Big Five Dimension is represented (Fig. 13).
3. The third subsection presents the user’s LinkedIn information in a basic format. Along with some basic information (name, industry, birthday and location of the user) the current and past working experience is provided, the user’s skills, education and languages spoken (Fig. 14).

4. The next subsection includes the user’s connections, presented along with their location and the industry they belong to (Fig. 15).

There is also the option to revoke the authorization of the application and, at last, the contact information, in case someone wants to get in contact with the developers.
5.2. Team Building Search Interface

In order for a team to be created there are several search criteria available, already explained in Section 4.4. They have all been implemented accordingly, as seen in Fig. 16.

As we can see, for the criteria “skills”, “languages” and “industries”, there is the option to match users to any, or all of the selected values. For example, if the language option is set to “all”, and the search values are English and Greek,
then we are searching for users who speak both languages. If it was set to “any”, we would search for users speaking either one, but not necessarily both. For each of the criteria that also work as a similarity measure (i.e. skills, languages, industries, location, education, and social connectivity), there is the option to turn it on or off. This does not affect the search procedure, and if any of the criteria are set, the results are matched accordingly. It means that the similarity of the teams is calculated based on the active (turned on) criteria, while those who are turned off are not considered in the calculations.

Concerning the personality criterion, we see, that it is broken down to the Big Five personality dimensions (Openness to Experience, Conscientiousness, Extraversion, Agreeableness, Neuroticism). Each dimension can be defined by a relative value (low, medium, high) or be turned off entirely. This can be done with the on/off option, that each dimension has available. For example, a search may be made, where the users should have high values of Openness to Experience, low values of Neuroticism, while the three remaining dimensions remain turned off, and thus are not taken into consideration (Fig. 17).

After all the appropriate options are set, a search in the database is initiated. The results page is divided into two subsections. In Fig. 18 a sample results page is displayed. Names and pictures have been removed from the preview.

On the right subsection, the individual users, who have been matched against the search criteria, are presented in a list. On the left subsection, all possible teams are presented. Each team contains its ranking number and similarity percentage, as well as each member’s picture and name. With a mouseover on the similarity percentage, an information balloon pop-ups containing each individual similarity of all the selected similarity measures.
Fig. 18: Results page. The first 3 results are displayed, out of 84 total.

In Fig. 19 an example is presented, where all similarity measures have been selected, and in Fig. 20 another example, where two similarity measures (languages, education) have been disabled and therefore are not taken into consideration during the calculations.
There is also the option to re-rank the teams based on one of the active similarity measures. An example is given in Fig. 21, where the re-ranking was applied, based on the educational similarity measure. The result may seem odd, as team #2 has nearly half the value of team #3 in their general similarity, but team #2 has 100% educational similarity, while team #3 follows with 80%.
6

Experimentation and Discussion

So as to assess the correctness and efficiency of the team building method we created (as described in Section 4), we provide a series of scenarios, namely a series of hypothetical searches towards our database. These scenarios try to highlight the features that are available, the way they influence the results, and the manner in which results can be adjusted.

6.1. Search Scenarios

In late March 2013 we asked a group of people to use our application, submit their data, and also fill in the personality questionnaire. We made a second call in late August 2013. Unfortunately, only 12 people shared their LinkedIn information in total, and just 9 of them completed the personality questionnaire.

For each one of the following scenarios we present a small and basic description, the search criteria that are selected and set and the reason behind this selection, the criteria which are active for the similarity calculations between team members, as well as the number of search results returned, the teams created, their final similarity value and the proposed team. In all examples, user anonymity is applied, and names have been removed. Nonetheless, the new symbolic names retain the original user associations within each team.
6.1.1. Scenario #1

**Description:** A 3-member team is requested, with a background in Higher Education. The team should have the optimal personality composition, in accordance with the corresponding research. Extraversion, for this scenario, is irrelevant.

**Search Criteria:** According to personality related research, as analyzed in Section 2, the two dimensions mostly correlated with team efficiency are Conscientiousness and Agreeableness. The lack of Neuroticism seems to be a positive factor, while Openness to Experience hasn’t so far indicated any relation whatsoever. Extraversion could play a positive role, but for this example we focus on the first three dimensions. As such, we set the following search criteria:

- Number of people: 3
- Personality dimensions:
  - Conscientiousness level: high
  - Agreeableness level: high
  - Neuroticism level: low
  - Openness to Experience: disabled
  - Extraversion: disabled
- Industry: Higher Education

**Similarity based on:** All criteria

**Individual matching results:** 4

**Proposed teams:** 4; see Table 4 for details.

**Result:** The proposed team has a total similarity value of 62.04%. We note, however, that the skill similarity of each team is relatively low, with a maximum of 30.77% for team #3. If the particular similarity measure is deemed irrelevant to the objective, it could be disabled. In that case, the first team has a final similarity value of 74.34%
### Table 4: Proposed teams for Scenario #1

<table>
<thead>
<tr>
<th>Team #1</th>
<th>User A</th>
<th>User B</th>
<th>User D</th>
<th>skills: 12.82%</th>
<th>languages: 100%</th>
<th>industries: 52.04%</th>
<th>location: 100%</th>
<th>education: 33.33%</th>
<th>social connectivity: 83.33%</th>
</tr>
</thead>
<tbody>
<tr>
<td>62.04%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team #2</td>
<td>User B</td>
<td>User C</td>
<td>User D</td>
<td>skills: 10.26%</td>
<td>languages: 97.78%</td>
<td>industries: 96.11%</td>
<td>location: 100%</td>
<td>education: 23.33%</td>
<td>social connectivity: 56.30%</td>
</tr>
<tr>
<td>56.3%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team #3</td>
<td>User A</td>
<td>User C</td>
<td>User D</td>
<td>skills: 30.77%</td>
<td>languages: 100%</td>
<td>industries: 52.04%</td>
<td>location: 100%</td>
<td>education: 23.33%</td>
<td>social connectivity: 52.46%</td>
</tr>
<tr>
<td>52.46%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team #4</td>
<td>User A</td>
<td>User B</td>
<td>User C</td>
<td>skills: 7.69%</td>
<td>languages: 97.78%</td>
<td>industries: 52.04%</td>
<td>location: 100%</td>
<td>education: 33.33%</td>
<td>social connectivity: 25%</td>
</tr>
<tr>
<td>49.01%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### 6.1.2. Scenario #2

**Description**: A 2-member team for a PR event concerning Thessaloniki, taking place in the city with an international audience. The team should work in unison and get along effortlessly.

**Search Criteria**: A way to ensure that the members of the team will get along well is to choose two people that may have a previous acquaintance or even be friends, rather than choose two complete strangers. This can be
achieved, by setting the social connectivity between members of each team strictly to 1st degree. The position is in the realm of PR, and as such they should be extroverted and talkative, which are traits of Extraversion. The event expects an international audience so they should also be open-minded, quick-witted and imaginative, all traits of Openness to Experience. As it concerns the city of Thessaloniki, candidates should be local and speak both Greek and English. As a result, we set the following search criteria:

- Number of people: 2
- Personality dimensions:
  - Extraversion level: high
  - Openness to Experience level: high
  - Conscientiousness: disabled
  - Extraversion: disabled
  - Neuroticism: disabled
- Languages: English, Greek
- Location: Greece
- Social Connectivity: 1st degree

**Similarity based on:** All criteria, but we could limit them to languages, location, social connectivity, and maybe education, as they could be viewed as more relevant.

**Individual matching results:** 2

**Proposed teams:** 1; presented in Table 5.

**Result:** Only one two-member team can be proposed, with a final similarity value of 59.5%. If we calculate similarity alternatively, as described above, the final similarity is 76.92%.

---

5 For Greece, as well as other countries, LinkedIn does not offer further specification of City/Region within the country. If it was possible, we would have set the city sub-option to Thessaloniki.
### Table 5: Proposed team for Scenario #2

<table>
<thead>
<tr>
<th>Team #1</th>
<th>User A</th>
<th>User B</th>
<th>skills: 25%</th>
<th>languages: 100%</th>
<th>industries: 30%</th>
<th>geographical: 100%</th>
<th>educational: 0%</th>
<th>social connectivity: 100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>59.5%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 6.1.3. Scenario #3

**Description**: A job posting asks for post-graduate web programmers with database background, i.e. with knowledge of MySQL, PHP and JavaScript, for pair programming.

**Search Criteria**: In this case the pre-requisites are almost explicitly stated. We could infer, in addition, that the industry the users should belong to is Internet or Computer Software. Therefore, we set the following search criteria:

- Number of people: 2
- Skills: MySQL, PHP, JavaScript
- Education: Master or greater
- Industry: Internet, Computer Software

**Similarity based on**: All criteria

**Individual matching results**: 3

**Proposed teams**: 3; see Table 6 for details.

**Result**: The proposed team has a more than sufficient similarity value of 80.76%, with a significant 15% difference from the next possible team.
Table 6: Proposed teams for Scenario #3

<table>
<thead>
<tr>
<th>Team #1</th>
<th>User A</th>
<th>User C</th>
</tr>
</thead>
<tbody>
<tr>
<td>80.76%</td>
<td>skills: 26.32%</td>
<td>languages: 100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>industries: 100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>geographical: 100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>educational: 70%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>social connectivity: 100%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Team #2</th>
<th>User A</th>
<th>User B</th>
</tr>
</thead>
<tbody>
<tr>
<td>65.86%</td>
<td>skills: 38.46%</td>
<td>languages: 93.33%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>industries: 94.44%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>geographical: 100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>educational: 100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>social connectivity: 0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Team #3</th>
<th>User B</th>
<th>User C</th>
</tr>
</thead>
<tbody>
<tr>
<td>61.07%</td>
<td>skills: 30.77%</td>
<td>languages: 100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>industries: 96.11%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>geographical: 100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>educational: 70%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>social connectivity: 0%</td>
</tr>
</tbody>
</table>

6.2. Discussion

The biggest challenge we encountered was user participation, which was critically low. This may be due to three reasons. First and foremost, not everyone uses LinkedIn as their social network of choice. This coincides with the fact that our major pool of potential users resides in Greece, where LinkedIn is not quite popular: Greece ranks 51st out of 87 countries measured, with 683,987 users in total, with a 6.22% population penetration⁶. Secondly, our call for participation was within the academic boundaries. As Skeels & Grudin (2009) have observed, LinkedIn is not really targeted to students, and the majority of its users are young professionals. The third reason could lie in the innate nature of people to be sceptical when it comes to sharing their personal information, and consequently decided against participating.

User participation is crucial when it comes to experiments. For us, the optimal number of participating users would be a few dozen, which would suffice to make more diverse use case scenarios and interesting experiments. A possible experiment could be the following: For a university course, participating students are usually given assignments to complete, either theoretical, technical or a combination thereof. In an assignment designed to be completed by a group of students (maybe of up to 4 people), we could set our application into action. After students have formed their teams, we can replicate them with our application. Then, after the assignment is complete and grades are announced, we can cross reference each team’s similarity value with this grade. The comparison can indicate if there is indeed a relation between team similarity and the team performance in the given assignment. It can also allow us to reconfigure or fine tune the similarity measures according to these results. The success of such an experiment, however, depends mostly on user participation, which unfortunately is tepid at best.

The personality questionnaire was completed by 75% of participating users. This value is quite sufficient, but at the same time somewhat unrepresentative. User participation is low, and as such, in case it increases we can’t be sure that this percentage will remain unchanged, or if it will start decreasing. As already explained in Section 4.3, we created a personality questionnaire consisting of 50 items, so that it can have some merit and not be too coarse-grained. Users who did not complete it, could have postponed it due to time limitations at that moment and never got round to it later.

Another small hurdle is the quality and completeness of the data that users provide through their LinkedIn accounts. When creating their LinkedIn profiles, some people prefer to fill only the fields that they deem necessary and leave the rest blank. By looking through our database, we have found distinct users who have left different fields blank, for instance their education, their skills, or the languages they speak. For example, for the proposed team of Scenario #2, the educational similarity of the two users is 0% (as seen in Table 5). By looking up the database information, we find out that one of the two users has not provided information concerning his education. As a result, the similarity of the two users is zero. An amendment to our logic could be to disregard the similarity for a particular category in cases like this, where there is no information provided from one or both users. However, regardless of the path we may choose to follow, there is a lurking risk, where, either by removing or by nullifying a specific similarity measure, we may be left in the end with too few.
When it comes to data usage, LinkedIn is quite strict in its policy. Individual user data acquired through the API must be presented to that particular individual and not be shared with other users or third parties. Furthermore, LinkedIn gives us access only to a user’s immediate connections. In this way, we cannot form a broader understanding of the intricate relationships between users. Hence, we know a user’s 1st degree connections and we can infer some of their 2nd degree connections, through other users that we have in our database. In no case can we know a user’s 3rd degree connections. Had we access to more information, we could have formed a connectivity graph. Then we could compute a user’s closeness, degree, or betweenness centrality and provide a more robust similarity measure concerning the social connectivity of the users.

A final topic worth mentioning concerns the available literature on personality and team dynamics. As LePine, et al. (2011) observe, there is a certain link between team member personality and team functioning and effectiveness, but the difficulty lies in the fact that “we have yet to develop an appreciation of which specific associations among specific traits and criteria are most important and useful with regards to human resource management decisions”. This means that despite the many hypotheses researchers have proposed about what personality traits are more suitable for the efficiency and cohesion either of every team, or of teams that belong to a specific field (e.g. in managements, sales, etc.), there is still work to be done, for them to be completely and accurately verified. Nevertheless, there are of course quite a few solid results that we have already examined (as presented in Section 2) and employed in our application, and as more research becomes available we will be able to update it accordingly.
Conclusion and Future Work

In order for teams to become more efficient, research has indicated alternative ways for their creation. One of them is through personality matching of team members. We devised a new method for creating teams, which combines users’ personality characteristics and their professional profile. We use a personality questionnaire to discern each user’s personality, utilizing the Big Five Factor Model, according to which personality can be described by five major aspects. These are: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. For their professional profile, we draw information from their LinkedIn accounts.

The creation of any team starts with setting the appropriate options. The available criteria for our version of team building are: Team size, users’ skills and spoken languages, their geographic location, their educational background, the industry they belong to, as well as their social connectivity to each other and, of course, their personality characteristics. After any of these criteria is set, a search in our database fetches the corresponding users. Then, all possible teams are created and we calculate the similarity between each team’s members. In order to do so, we have created a number of similarity measures, corresponding to all search criteria, apart from the personality one. The next step is to check if each team’s group personality matches the one that has been defined, and remove those which do not. In the end, the teams are sorted and presented in a descending order.

In order to test our method we invited users to share their professional profile data and fill in the provided personality questionnaire. During this procedure, we faced a number of limitations. The primary caveat was user participation, which
was exceedingly low, a fact that did not allow us to make the experiments we
would have liked in their full extent. Secondly, the data users provided were at
times deficient, a result of their incomplete LinkedIn accounts. For example,
depending on the user, a particular section was missing, such as the skill, language
or education one. Likewise, a small number of users (25%) did not fill in the
personality questionnaire. Another limitation is imposed by the LinkedIn policy,
which does not give us access to a more general view of the user connection
network. We only know each user’s immediate connections, and through our
database we can only deduce a smaller and incomplete part of the network.

Finally, as personality is not a strictly quantifiable entity, we found it rather
difficult to translate the available research on team dynamics and personality to
exact numbers. Furthermore, research with solid results is somewhat sparse, and
therefore our sources were limited. Personality based team building is,
nonetheless, an important step forward, as far as team efficiency is concerned, and
we believe that we have set the groundwork and created a solid framework
towards this direction. We hope that more relevant research will become available
in the future, so as to improve the personality feature of our team
recommendation algorithm.

As a next step, our intention is to conduct more experiments in order to better
assess our team building method. Come this fall, we would like to initiate an
experiment like the one described in Section 6.2 and this time, so as to overcome
the low participation issue, strongly prompting students to take part in it and
helping us. This could be done either in a postgraduate or an undergraduate
course. For the former, the students are older, more mature and with a more
defined skill set. It is also, more probable that they will have a LinkedIn account.
The main merit of the latter is that the pool of potential user is larger and even if
participation turns out to be low, it could account to more than all of the
postgraduate student capacity. We would also like to delve even deeper into the
available literature so as to create more team personality profiles and provide
ready templates for team creation (for example for specific areas –professional,
technical, academic, etc–, or even specific professions).
References


Hathaway, S. R. & McKinley, J. C., 1951. *Minnesota Multiphasic Personality Inventory; manual (Revised)*, s.l.: s.n.


Appendix

9.1. IPIP Personality Questionnaire

Users have to fill out the 50-item questionnaire of IPIP Big Five Factor Markers. Every sentence begins with “I…” referring to the common attitude the user has towards it. The user can agree, strongly agree, disagree, strongly disagree or be neutral with each sentence.

1. Am not easily bothered by things.
2. Make friends easily.
3. Do just enough work to get by.
4. Am the life of the party.
5. Carry the conversation to a higher level.
6. Have little to say.
7. Tend to vote for liberal political candidates.
8. Make people feel at ease.
10. Avoid philosophical discussions.
11. Suspect hidden motives in others.
12. Rarely get irritated.
13. Pay attention to details.
14. Do not like art.
15. Dislike myself.
16. Enjoy hearing new ideas.
17. Find it difficult to get down to work.
18. Do not enjoy going to art museums.
19. Remain calm under pressure.
20. Waste my time.
21. Have a good word for everyone.
22. Am very pleased with myself.
23. Don't like to draw attention to myself.
24. Get back at others.
25. Make plans and stick to them.
26. Have a vivid imagination.
27. Panic easily.
28. Don't talk a lot.
29. Get chores done right away.
30. Would describe my experiences as somewhat dull.
31. Have a sharp tongue.
32. Keep in the background.
33. Insult people.
34. Am skilled in handling social situations.
35. Am often down in the dumps.
36. Believe that others have good intentions.
37. Shirk my duties.
38. Am not interested in abstract ideas.
39. Cut others to pieces.
40. Know how to captivate people.
41. Carry out my plans.
42. Have frequent mood swings.
43. Feel comfortable with myself.
44. Accept people as they are.
45. Tend to vote for conservative political candidates.
46. Don't see things through.
47. Feel comfortable around people.
48. Respect others.
49. Believe in the importance of art.
50. Often feel blue.
9.2. Probabilities and Rewards of Industries and their corresponding Groups for the Industry similarity measure

In Probability Theory an experiment is an action with an uncertain result, such as a coin or die toss. A possible result of an experiment is called an outcome ($\omega$). For example, the outcome of a coin toss could be either “heads” or “tails”. The set of outcomes of each experiment is the sample space ($\Omega$). In the example of the die toss, the sample space is $\{1, 2, 3, 4, 5, 6\}$. An event ($A$) is a single result of an experiment ($A \subseteq \Omega$) and can consist of multiple outcomes. For example, the event of “rolling an odd number” is $\{1, 3, 5\}$.

The probability of an event is defined as the ratio of the number of all desired outcomes to the number of all possible outcomes (i.e. the sample space):

$$P(A) = \frac{|A|}{|\Omega|}$$

So, the probability of the event $A$ to “roll an odd number” is:

$$P(A) = \frac{|\{1, 3, 5\}|}{|\{1, 2, 3, 4, 5, 6\}|} = \frac{3}{6} = 50\%$$

As described in Section 4.5.3, each user can belong to an industry. According to LinkedIn there are 147 Industries, which, can belong to 1, 2, or 3 out of 17 possible general groups. These groups are presented in Table 7. If two users have the exact same industries the score is obviously 1, while on the opposite occasion, the score between two users’ industries is defined as:

$$\text{overlap coefficient} \cdot (1 - \text{probability(event)})$$

$\text{probability(event)}$ refers to the probability of the two industries belonging in their respective groups.

We remind that:

$$\text{overlap coefficient}(A, B) = \frac{|A \cap B|}{\min \left(|A|, |B|\right)}$$
There are 10 distinct cases, which are described below. Each case describes the relation between two industries. For each pair, we provide the number of common groups, the number of total groups for each industry, their overlap coefficient, the probability for that particular event, as well as the total score. For example, we can have one industry belonging to one group, and one industry belonging to two groups. The first can have 1 out of 1 group same \( \left( \frac{1}{1} \right) \) and the second 1 out of 2 \( \left( \frac{1}{2} \right) \).

We also remind that:

\[
\text{combinations}(n,r) = \frac{n!}{(n-r)! \cdot r!}
\]

1. **A: 1/1, B: 1/1**

\[
\text{overlap coefficient} = \frac{1}{1} = 1
\]

\[
P(A \cap B) = \frac{17}{17 \cdot 17} = \frac{1}{17}
\]

The sample space consists of \( 17 \cdot 17 \) possible outcomes. For 17 of them the group of each industry is the same.

\[
score = 1 \cdot \frac{16}{17} = 0.9412
\]
2. A: 1/1, B: 1/2 (or A: 1/2, B: 1/1)

\[
\text{overlap coefficient} = \frac{1}{1} = 1
\]

\[
P(A \cap B) = \frac{16}{136} = \frac{2}{17}
\]

In order to find the available pairs out of the 17 available groups, we calculate the combinations, where order is not taken into consideration and repetition is not allowed:

\[
\text{combinations}(17, 2) = \frac{17!}{15! \cdot 2!} = 136
\]

The first industry could belong to any of the 17 groups. The second industry belongs to two groups, and either one must match the group of the first industry. From 136 pairs, each distinct group occurs in 16 of them (i.e. the combination of one group with the rest).

\[
\text{score} = 1 \cdot \frac{15}{17} = 0.8824
\]

3. A: 1/2, B: 1/2

\[
\text{overlap coefficient} = \frac{1}{2} = 0.5
\]

\[
P(A \cap B) = \frac{136 \cdot 30}{136 \cdot 136} = \frac{30}{136} = \frac{15}{68}
\]

In this case, the sample space is:

\[
\Omega = 136 \cdot 136
\]

The first industry can belong to any of the 136 available group pairs (Industry A \( \in \{i,j\} \)). The second industry must belong to a group pair, so that they have only one group in common. For Industry B we assume that the first group it belongs is \( i \). In order to have a match, the second group can be any of the remaining 16, apart from \( j \), because in that case we would have a 2/2
match. The reverse applies for pairs containing \( j \). So, for every group pair

*Industry A* may belong to, there are \( 2 \cdot 15 = 30 \) possible pairs for *Industry B*.

\[
\text{score} = 0.5 \cdot \frac{53}{68} = 0.3897
\]

4. A: 2/2, B: 2/2

overlap coefficient \( \frac{2}{2} = 1 \)

\[
P(A \cap B) = \frac{136}{136 \cdot 136} = \frac{1}{136}
\]

There are 136 distinct group pairs for *Industry A* to belong to. *Industry B* must belong to the exact same pair, out of the \( 136 \cdot 136 \) total pairs, that compose the sample space.

\[
\text{score} = 1 \cdot \frac{135}{136} = 0.9926
\]

5. A: 1/1, B: 1/3 (or A: 1/3, B: 1/1)

overlap coefficient \( \frac{1}{1} = 1 \)

\[
P(A \cap B) = \frac{17 \cdot 120}{17 \cdot 680} = \frac{3}{17}
\]

Out of 17 available groups we can create 680 triples.

\[
\text{combinations}(17, 3) = \frac{17!}{14! \cdot 3!} = 680
\]

*Industry A* may belong to one of the 17 groups. Out of the 3 groups of *Industry B*, one must match that of *Industry A*. Then, the two other groups can be one of the remaining 16. All possible combinations of this occasion are:

\[
\text{combinations}(16, 2) = 120
\]

So, there are \( 17 \cdot 120 \) matching combinations for *Industries A and B*.

\[
\text{score} = 1 \cdot \frac{14}{17} = 0.8235
\]
6. A: 1/2, B: 1/3 (or A: 1/3, B: 1/2)

\[
\text{overlap coefficient} = \frac{1}{2} = 0.5
\]

\[
P(A \cap B) = \frac{136 \cdot 2 \cdot 105}{136 \cdot 680} = \frac{210}{680} = \frac{21}{68}
\]

The sample space \( \Omega \) is:

\[
c(17,2) \cdot c(17,3) = 136 \cdot 680
\]

For every pair of groups \( i,j \), that Industry A may belong to, the probability is 1/136. We only want to match one group. So, out of the three possible groups for Industry B, we assume that one is \( i \). Then, the remaining two, cannot be \( j \), so we are left with 15 possible group combinations: combinations \( (15,2) = 105 \). The same applies for \( j \).

\[
\text{score} = 0.5 \cdot \frac{47}{68} = 0.3456
\]

7. A: 2/2, B: 2/3 (or A: 2/3, B: 2/2)

\[
\text{overlap coefficient} = \frac{2}{2} = 1
\]

\[
P(A \cap B) = \frac{136 \cdot 15}{136 \cdot 680} = \frac{15}{680}
\]

For every pair of groups \( i,j \), that \textit{Industry A} may belong to, the probability is 1/136. In this case we want to match both groups with 2 out of 3 groups of \textit{Industry B}. As such, we assume that the first two groups are indeed \( i \) and \( j \). The third group can be one of the remaining 15.

\[
\text{score} = 1 \cdot \frac{665}{680} = 0.9779
\]
8. A: 1/3, B: 1/3

\[
\text{overlap coefficient} = \frac{1}{3} = 0.33
\]

\[
P(A \cap B) = \frac{680 \cdot 3 \cdot 91}{680 \cdot 680} = \frac{273}{680}
\]

Suppose the three groups of Industry A are \(i, j, k\). There are 680 possible outcomes for this event. In order to have only one in common with Industry A, Industry B should belong to \(i\), but not \(j\) and \(k\), to \(j\), but not \(i\) and \(k\), or to \(k\), but not to \(i\) and \(j\). For each of these three events, the outcome for the two remaining groups is 91:

\[
\text{combinations (14, 2)} = 91
\]

Finally:

\[
\text{score} = 0.33 \cdot \frac{407}{680} = 0.1995
\]

9. A: 2/3, B: 2/3

\[
\text{overlap coefficient} = \frac{2}{3} = 0.66
\]

\[
P(A \cap B) = \frac{680 \cdot 3 \cdot 14}{680 \cdot 680} = \frac{42}{680} = \frac{21}{340}
\]

Suppose the three groups of Industry A are \(i, j, k\). In order to have two groups in common, Industry B should belong to \(i\) and \(j\), but not \(k\), to \(i\) and \(k\), but not \(j\), or to \(j\) and \(k\), but not \(i\). For each of these events, the outcome of the remaining group is 14. So:

\[
\text{score} = 0.66 \cdot \frac{319}{340} = 0.6255
\]
There are 680 distinct triples. As a result, there are 680 outcomes, where both Industries belong to the same three groups.

\[
\text{overlap coefficient} = \frac{3}{3} = 1
\]

\[
P(A \cap B) = \frac{680 \cdot 1}{680 \cdot 680} = \frac{1}{680}
\]

\[
\text{score} = 1 \cdot \frac{679}{680} = 0.9985
\]