

This is a post-refereeing version of a manuscript published by ASCE.

Please, in order to cite this paper properly:

**Ballesteros-Pérez, P., Phua, F.T.T., Mora-Melià, D. (2018) “Human Resource Allocation to Multiple Projects Based on Members’ Expertise, Group Heterogeneity, and Social Cohesion.” *Journal of Construction Engineering and Management*, 145(2), 04018134.**

[https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001612](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001612)

The authors recommend going to the publisher’s website in order to access the full paper.

If this paper helped you somehow in your research, feel free to cite it.

1                   **Human resource allocation to multiple projects based on**  
2                   **members' expertise, group heterogeneity and social cohesion**

3 Pablo Ballesteros-Pérez, Ph.D.<sup>1\*</sup>; Florence Ting Ting Phua, Ph.D.<sup>2</sup>; Daniel Mora-Melià, Ph.D.<sup>3</sup>

4   **Abstract**

5 Project managers regularly allocate human resources to construction projects. This critical  
6 task is usually executed by fulfilling the minimum project staffing requirements normally  
7 based around the quantity and competence of project members. However, research has shown  
8 that team performance can increase by up to 10% and 18%, respectively, as a consequence of  
9 the group members' heterogeneity and social cohesion. Also, there is currently no practical  
10 quantitative tool which incorporates these aspects to allow project managers to achieve this  
11 task efficiently and objectively.

12 A new quantitative model for the effective allocation of human resources to multiple projects,  
13 which takes into account group heterogeneity and social cohesion is proposed. This model is  
14 easy to build, update and use in real project environments with the use of a spreadsheet and a  
15 basic optimization engine (e.g. Excel Solver). A case study is proposed and solved with a  
16 Genetic Algorithm to illustrate the model implementation. Finally, a validation example is  
17 provided to exemplify how group heterogeneity and social cohesion condition academic  
18 achievement in an academic setting.

---

<sup>1</sup> Senior Lecturer. School of Architecture, Building and Civil Eng., Loughborough University, Loughborough,  
Leicestershire LE11 3TU, UK. \*Corresponding author: +44 (0) 150 922 3771, [p.ballesteros-perez@lboro.ac.uk](mailto:p.ballesteros-perez@lboro.ac.uk)

<sup>2</sup> Associate Professor. School of the Built Environment. University of Reading,  
Whiteknights, Reading RG6 6DF, UK. Phone: +44 (0) 118 378 8194, [f.phua@reading.ac.uk](mailto:f.phua@reading.ac.uk)

<sup>3</sup> Associate Professor. Departamento de Ingeniería y Gestión de la Construcción, Universidad de Talca, Facultad  
de Ingeniería, Curicó, camino Los Niches km.1 3340000, Chile. Phone: +56 951 265 044, [damora@utalca.cl](mailto:damora@utalca.cl)

19

## 20 **Keywords**

21 Human resource management; human resource allocation; team performance; heterogeneity;  
22 faultlines; cohesion; sociometry

23

## 24 **Introduction**

25 It is a well-known saying that ‘people are the lifeblood of organizations’. Indeed,  
26 despite living in an era of constant technological advancement, most of our tasks are still  
27 done, handled or supervised by human beings. In organization life, the size and/or complexity  
28 of many undertakings nowadays demand the involvement of many people (sometimes from  
29 different organizations) working together to achieve a common goal. This goal can be  
30 anything, but many times involves creating deliverables (products, services) to enhance a  
31 company’s internal performance, to make profit, or both. However, people (employees,  
32 workers) who take part in these undertakings are normally subject to constraints. For  
33 example, they are qualified to do certain jobs and not others; they have different levels of  
34 competence in different domains; they cannot be present in multiple locations; and, certainly,  
35 they have physical constraints in terms of how long they can work for (Hendriks et al. 1999).

36 Therefore, when there are several, sometimes concurrent projects that require the  
37 participation of people to be completed, a project manager faces a practical dilemma: how to  
38 best allocate his/her human resources on-hand to deliver his/her projects successfully.  
39 ‘Successfully’ can mean completing the projects on time, on budget and within an agreed (or  
40 shared) quality threshold, or just meeting the key stakeholders’ expectations (Xia et al. 2017).  
41 In any case, as long as there are ongoing projects, the project manager will require competent  
42 human resources to engage in certain tasks for a period of time before they are freed and able

43 to join other ongoing or upcoming projects. An essential part of the project manager's role in  
44 the allocation of optimum human resources is to ensure, as much as possible, that the  
45 individuals within the projects can work cooperatively with each other (Anvuur and  
46 Kumaraswamy 2016).

47         However, collaboration between project members does not happen by chance. There  
48 are indeed many factors that prevent this from happening. These factors can be  
49 communication-related for instance, and/or have to do with the project member's  
50 demographic attributes such as (differences in) nationality, education, religion, experience, to  
51 cite a few (Al-Bayati et al. 2017). Sometimes, there are people who do not like working with  
52 certain individuals, and this can also be really detrimental to the project progress and its  
53 eventual success (Chen et al. 2017b). In this regard, Phua (2004) and Phua and Rowlinson  
54 (2004) have found that cooperative behavior between project members is influenced, to a  
55 certain extent by individual members' intrinsic social and psychological factors which have  
56 to do with many more factors other than just their extrinsic demographic profile such as age,  
57 sex, education, work experience and roles. For this reason, we will consider both cohesion  
58 and heterogeneity factors later when aiming to build high-performing teams.

59         Given our existing understanding of the various factors that affect team performance,  
60 there is however, a scarcity of quantitative and objective tools that enable the effective  
61 allocation of human resources in terms of where and when they are to be allocated to projects  
62 (Ahmadian Fard Fini et al. 2017). Conventionally, this type of allocation issues largely fall  
63 within mainstream Human Resource Management (HRM) application which has its roots in  
64 social sciences.

65         A different, maybe opposite, scenario can be found within Operational Research  
66 (OR), which deals with the modeling and application of advanced analytical methods to make  
67 better decisions. The problem of allocating multiple human resources to a single project is

68 relatively recent in OR, but it has been well studied and is known nowadays as the ‘Team  
69 Formation Problem’ (TFP) (Tseng et al. 2004). When there are multiple simultaneous  
70 projects, the TFP becomes the ‘Multiple Team Formation Problem’ (MTFP). Particularly, the  
71 grouping of individuals to create teams have been made by attending to multiple factors: the  
72 resources’ temporal availability, current workload, individuals’ skills, level of competence,  
73 geographical distance, seniority, number of contacts, among many others (Gutiérrez et al.,  
74 2016). In this line of research, it is not common to find theoretically-grounded sociological  
75 considerations in the composition of teams. This means that, whereas it is relatively easy to  
76 come across OR models that allocate resources that meet some functional (e.g. skills,  
77 competence) project members’ requirements, it is very rare to find models that try to optimize  
78 other socially-based group traits like intra-group social preferences and group cohesion  
79 (Ballesteros-Pérez et al. 2012). This piece of research proposes to take a step forward in  
80 bridging this gap.

81         In this paper, a new human resource allocation model that takes into account, not just  
82 basic project staff requirements and employees’ profiles, but also group heterogeneity  
83 (diversity) and social cohesion, is developed. This is a worthwhile contribution because, as  
84 discussed earlier, team performance has been demonstrated to be significantly influenced by  
85 these two factors. Hence, it seems logical to incorporate this knowledge when creating high-  
86 functioning teams which comprise the ‘right’ individuals working together. To this end, the  
87 rest of the paper will be structured as follows. The *literature review* section will go over the  
88 major contributions published in the areas of the MTFP, group heterogeneity and social  
89 cohesion. The *materials and methods* section will formulate the model, define its major  
90 variables and explain how these are interrelated under mathematical expressions for  
91 measuring team performance. A *case study* will exemplify the model implementation in a  
92 fictitious company environment with twenty people and three simultaneous projects. A short

93 *validation* section will implement the model in a real academic setting where a cohort of 15  
94 MSc students worked in groups to deliver three projects. The *discussions* will provide some  
95 insight and further analysis on the implications and limitations of the model. Finally, the  
96 *conclusions* will summarize the paper and convey why the proposed tool is relevant to the  
97 wider project management community.

98

## 99 **Literature review**

100 The proposed model draws from research developed in two very different areas –  
101 operational research (OR) and applied psychology (AP) –, but it is applied on a third one:  
102 Human Resource Management (HRM). The amount of works published in connection with  
103 HRM within both OR and AP is endless, so it is necessary to narrow down significantly the  
104 works to be presented here. In this regard, only three very relevant topics will be reviewed:  
105 the MTFP, group heterogeneity and faultlines, and group cohesion and sociometry.

106

### 107 ***The Multiple Team Formation Problem (MTFP)***

108 The MTFP involves the distribution of people with different skillsets to a series of  
109 teams (projects) that usually require more than a single area of expertise while optimizing  
110 other criteria (e.g. profits, execution time, number of people). This problem is known to be  
111 NP-hard (Non-deterministic Polynomial-time Hard) even for instances with a single project  
112 (Gutiérrez et al., 2016). This means the MTFP belongs to the set of OR problems that are  
113 harder to solve.

114 The first attempt to model and compute a solution to the TFP is relatively recent and  
115 was developed by Lappas et al. (2009) when trying to create teams of experts from  
116 professional profiles posted on social networks. Just a year later, Dorn and Dustdar (2010)

117 proposed solving the TFP with a first heuristic approach, whereas Li and Shan (2010)  
118 improved the Enhanced-Steiner algorithm that was one of the two original algorithms used to  
119 solve the TFP.

120 A year later, Yin et al., (2011) were the first to consider social influence among the  
121 teams of experts. Additionally, Farhadi et al. (2011) allowed for the possibility of different  
122 competence levels among the human resources, a generalization that will also be considered  
123 in our model.

124 In 2012, the number of works published on the TFP grew exponentially. Among the  
125 most relevant: Sorkhi et al. (2012) proposed a game theoretic approach to form and rank  
126 project teams; Farhadi et al. (2012a, 2012b) extended the second original algorithm that had  
127 proven to be very effective when dealing with the TFP – the Rarest First algorithm –;  
128 whereas Gajewar and Sarma (2012) proposed three new optimization algorithms and  
129 successfully applied them to the MTFP for the very first time.

130 Next, Shi and Hao (2013) formulated the MTFP with a multi-criteria decision-making  
131 ranking approach involving the individuals' social networks. Then, Teixeira and Huzita  
132 (2014) approached the MTFP considering the human resources' contextual information  
133 (culture, idiom, temporal distance and previous experience), besides task requirements and  
134 the interpersonal relationships among human resources. Our proposed model will also take  
135 advantage of similar constructs in order to create a multi-dimensional model. Also, Agrawal  
136 et al. (2014) focused on educational settings allowing the MTFP to be implemented without  
137 allowing overlaps between the different student teams, a feature that will also be considered  
138 in our model. Still in the same year, Awal and Bharadwaj (2014) tried to capture the synergy  
139 produced among team members by means of a new ad-hoc concept named 'Collective  
140 Intelligence' and also used a Genetic algorithm to solve their problem formulation. In this  
141 paper, the solution of the case study proposed later will also make use of a genetic algorithm

142 approach as the way this ad-hoc index was defined share some similarities with our objective  
143 function.

144 Although there have been many other recent works published on the MTFP, these will  
145 not be recounted here as they are not directly germane to this study. However, one that is  
146 perhaps worth highlighting is the work from Gutiérrez et al. (2016) which formally included  
147 sociometric preferences among individuals in the MTFP. Our proposed model also shares a  
148 similar approach for modeling group cohesion. However, the algorithmic approach will be  
149 totally different to Gutiérrez et al.'s as our model includes other dimensions, which makes  
150 our model no longer quadratic.

151

### 152 ***Group heterogeneity and faultlines***

153 Research on how team effectiveness is influenced by the team composition has been  
154 abundant too. Most of this research has focused precisely on measuring and analyzing the  
155 effects of group heterogeneity on team performance. Group heterogeneity (homogeneity)  
156 refers to a measurement of how different (similar) the members' demographic attributes (age,  
157 sex, ethnicity, etc.) are with each other. There are many reviews on group heterogeneity (see  
158 Earley and Gibson (2002) for a comprehensive one) but they will not be recounted here  
159 either. In this piece of research, we are focusing on the quantitative aspects of how  
160 heterogeneity is measured and what are its effects on team performance, rather than the  
161 mechanisms or factors that cause it.

162 With this in mind, the first indices that captured quantitatively how diverse  
163 (homogeneous/heterogeneous) a group can be were defined by Blau (1977) and Allison  
164 (1978). Generally, these and other later indices involved measuring group homogeneity as the



165 members' demographic attribute overlaps. With those indices, heterogeneity was also  
166 generally defined as the inverse of homogeneity, that is  $\text{heterogeneity} = 1/\text{homogeneity}$ .

167         Additionally, for a long time, it was believed that the presence of faultlines  
168 (demographic features that divide a bigger group into two or more relatively homogeneous  
169 subgroups) was detrimental to group performance (Lau and Murnighan 2005). It was not  
170 until the work of Gibson and Vermeulen (2003), who proposed a new metric for measuring  
171 group heterogeneity – the Subgroup Strength – , that it was understood that the presence of  
172 subgroups (faultlines) could indeed promote team learning behavior and improve their  
173 performance. The Subgroup Strength (SS) has many advantages over previous homogeneity  
174 metrics (indices) as it allowed researchers to identify group faultlines much more effectively.  
175 Indeed, it was shown recently by Meyer and Glenz (2013) in a comprehensive comparative  
176 study that the SS is one of the simpler, yet more powerful metrics for measuring group  
177 heterogeneity in the presence of two or more subgroups. For these reasons, SS will also be  
178 used in our model later to describe subgroups' heterogeneity.

179         Finally, Gibson and Vermeulen (2003) also showed that a team's performance  
180 seemed to vary by up to 10% depending on the SS. Again, this was supported recently by  
181 another study by Chen et al. (2017). This study also confirmed another speculation of Gibson  
182 and Vermeulen's: that the relationship between SS and team performance was an inverted U-  
183 shape whose minima (lower performance) were to be expected for extremely homogeneous  
184 and heterogeneous groups. Finally, many other works have been published on the effects of  
185 group heterogeneity on intra- and cross-subgroups demographic faultlines (Lau and  
186 Murnighan 2005), but only some related to group cohesion will be reviewed later in the  
187 *Discussions* to clarify the effect of possible collinearities between both variables.

188

189        *Group cohesion and sociometry*

190            Group cohesion is a desirable attribute because research has proven it to be positively  
191 related to team performance, as well as a wide range of other positive behavioral outcomes  
192 (better individuals' attitude, well-being, lower absenteeism, etc.) (Chang and Bordia 2001;  
193 Chen et al. 2017b). However, very few pieces of research have actually quantified the extent  
194 to which team performance is influenced by group cohesion or dissociation.

195            One exception is a recent and comprehensive review performed by Evans and Dion  
196 (2012). These authors, beyond concluding that there is a positive relationship between  
197 cohesion and performance, recounted that cohesive groups seem to perform around 18  
198 percentile points on average above the average (uncohesive) groups. This figure will be used  
199 later in our model as other research has also corroborated the cohesion-performance  
200 relationship even when different settings (e.g. business, education, research) or group sizes  
201 are considered (Castaño et al. 2013). Furthermore, because existing research on cohesion and  
202 performance has operationalized cohesion almost completely in terms of interpersonal  
203 attraction (see evidence from Lott and Lott (1965) to Beal et al. (2003) for instance), it makes  
204 theoretical sense for our model to adopt sociometry to model group cohesion.

205            Sociometry was devised by Jacob Levy Moreno (Moreno 1941) and is a method that  
206 can be used for estimating the quality of group dynamics. It is one of the few methods that  
207 allows the gathering of quantitative information about the informal structure of a group that is  
208 difficult to obtain in other ways. Sociometry was extensively used between the 40s and 60s at  
209 schools, companies and research settings to examine social interrelations and communication  
210 patterns within groups (Salo 2006). In sociometry, interpersonal relations are measured by  
211 asking group members to express their preferences and rejections for particular companions  
212 in a certain situation or activity (Festinger et al. 1950). Hence, the advantageous simplicity of  
213 sociometry is, at the same time, its major limitation: it requires that group members are

214 truthful and open in stating who they prefer and not prefer to work with. A reasonable  
215 question then is whether group cohesion can be adequately represented by sociometric  
216 choices and if these choices can be eventually captured by means of questionnaires that  
217 request group members to state their preferences and rejections towards other group  
218 members. In fact, both aspects have been subjected to multiple research studies in many  
219 varied settings. An example of a brief but reassuring and confirmatory review can be found in  
220 Salo (2006).

221 Finally, there is one question that needs to be addressed before formulating the model.  
222 As stated earlier, the proposed model will group individuals under different projects that have  
223 some minimum staff (areas of expertise and levels of competence) requirements. According  
224 to a recent piece of research (Mathieu et al. 2015), when people with the right combination of  
225 expertise work together, as expected, this is positively related with team performance.  
226 However, this same piece of research also showed that this is unrelated to team cohesion.  
227 With this in mind, we will allow our model to effectively separate the effect of the constraints  
228 (i.e. minimum project staffing requirements) from the group performance variables (i.e.  
229 group heterogeneity and cohesion metrics).

230

## 231 **Materials and methods**

### 232 *Model outline*

233 In this section, an OR model that allocates a pool of skilled individuals to a series of  
234 simultaneous projects with specific staffing (expertise and competence) requirements is  
235 proposed and mathematically described in detail. This model will take into account how  
236 similar (homogeneous) these individuals are and how they get along with each other (group  
237 cohesion).

238

239 ***Mathematical notation***

240 Let us assume two individuals  $i$  and  $j$  where  $i, j$  belong to a set of  $n$  people (workers)  
241 who are available to be allocated into teams. Let us assume that these individuals can be  
242 combined into a number of non-overlapping teams (subgroups) where each team is noted by  
243 the letter  $k$  and whose size is noted as  $n_k$  (number of members of team  $k$ ).

244 For every individual  $i$  (or  $j$ ) it is assumed that the following information is known as  
245 illustrated in the following examples:

- 246 • Professional level of competence  $l_i$  where  $l_i \in L$  and  $L = \{\text{junior, intermediate, senior}\}$
- 247 • Functional department  $d_i$  where  $d_i \in D$  and  $D = \{\text{architecture, civil, mechanical, electrical}\}$
- 248 • Age  $a_i$  where  $a_i =$  positive integer.
- 249 • Gender  $g_i$  where  $g_i \in G$  and  $G = \{\text{Male, Female}\}$
- 250 • Ethnicity  $e_i$  where  $e_i \in E$ , and where  $E$ , for simplicity, will be assumed here as the  
251 continent of origin, that is  $E = \{\text{African, Antarctic, Asian, Australian, European, North}$   
252  $\text{American, South American}\}$
- 253 • Team tenure (seniority in the same group or company)  $t_i$  where  $t_i =$  positive integer.
- 254 • Sociometric preference of individual  $i$  towards individual  $j$ , that is  $s_{ij}$  where  $i \neq j$ ,  $s_{ij} \in S$  and  
255  $S = \{-1, 0, +1\}$ . Particularly,  $s_{ij} = -1$  means  $i$  dislikes working with  $j$ ,  $s_{ij} = 0$  means  $i$  is  
256 neutral towards (or has never worked with)  $j$ , and  $s_{ij} = +1$  means  $i$  likes working with  $j$ .
- 257 The set of all values  $s_{ij}$  correspond to a non-symmetrical matrix of size  $n \times n$ .

258 Sociometric preferences aside, these individuals' attributes have been selected here as  
259 they were the ones adopted by Gibson and Vermeulen (2003) in their seminal work on group  
260 faultlines. This set of attributes has been widely tested in subsequent research (e.g. Chen et

261 al. 2017a; Meyer and Glenz 2013) and it is still largely accepted that they provide a robust  
262 representation of group diversity.

263 Hence, given  $n$  people available from whom we know their  $l_i, d_i, a_i, g_i, e_i, t_i$  and  $s_{ij}$ , we  
264 will create subsets (subgroups/teams) of  $n_k$  individuals, each of which will be working on a  
265 different project  $k$ . Individuals can only be allocated to either a single group  $k$  or no subgroup  
266 at all (those unallocated individuals will be idle resources). This implies that no individual  
267 can be present in two or more subgroups, even if they could only work part-time in several  
268 projects. We use this simplified assumption to make this model more accessible from the  
269 point of view of its first mathematical formulation.

270 Therefore, as implied above, every subgroup  $k$  will be allocated to a single project and  
271 we will note projects and subgroups (teams) with the same subscript  $k$  from now on. Each  
272 project  $k$  will have specific staffing requirements ( $p_k$ ). For instance,  $p_k = \{1 \text{ senior Architect, 1}$   
273  $\text{intermediate civil engineer, 1 junior civil engineer, 2 intermediate electrical engineers}\}$ . Any  
274 subgroup of workers  $n_k$  that matches or exceeds (both in number and/or competence) these  
275 requirements will be considered a feasible subgroup that can potentially be allocated to  
276 project  $k$ .

277

### 278 ***Team performance measurement***

279 In order to determine which feasible allocation of subgroups is most desirable, it is  
280 necessary to anticipate how much better each possible alternative allocation of subgroups  
281 would perform if eventually chosen. Additionally, it is worth emphasizing that each feasible  
282 allocation might encompass multiple subgroups as each subgroup will be allocated to one  
283 project. Therefore, it is necessary to create an index that captures, not just how efficient each  
284 subgroup is, but also how efficient all groups are on average; that is, how efficient the

285 allocation is altogether. This index will be named ‘Global Efficiency (E)’ and will correspond  
 286 to a weighted average calculated from the subgroup Efficiencies of each subgroup  $k$  (noted as  
 287  $E_k$ ), that is:

$$288 \quad E = \sum E_k \cdot w_k \quad (1)$$

289 In this expression,  $w_k$  corresponds to the weight of each subgroup  $k$ . This way  $w_k$  can  
 290 be calculated, for instance, proportionally to each project  $k$ 's budget ( $b_k$ ). Alternatively,  $w_k$   
 291 can also be calculated proportionally to the number of people  $n_k$  from each project, divided  
 292 by the total people available  $n$  (allocated or not) or the total number of allocated people only  
 293 ( $\sum w_k$ ). These alternatives are expressed in equations (2) and (3), respectively:

$$294 \quad w_k = \frac{b_k}{\sum b_k} \quad (2)$$

$$295 \quad w_k = \frac{n_k}{\sum n_k} \quad \text{or} \quad w_k = \frac{n_k}{n} \quad (3)$$

296 With the global (allocation) Efficiency E defined in (1) as a function of each  
 297 subgroup's  $E_k$  and  $w_k$  values, now it is necessary to detail how  $E_k$  values can be calculated.

298  $E_k$  is a composite efficiency index obtained as the product of two other indices that  
 299 represent the expected performance of that subgroup  $k$  in terms of its homogeneity ( $P_k^{SS}$ ) and  
 300 social cohesion ( $P_k^S$ ). Namely,

$$301 \quad E_k = P_k^{SS} \cdot P_k^S \quad (4)$$

302 Particularly,  $P_k^{SS}$  estimates the Performance of a subgroup  $k$  based on the Subgroup  
 303 Strength (SS) as defined by Gibson and Vermeulen (2003). To calculate the SS value of a  
 304 subgroup  $k$  (noted as  $SS_k$ ), it will be necessary to calculate the subgroup  $k$ 's homogeneity  
 305 value  $h_k$  first, as well as the individuals' degree of overlaps in terms of different diversity

306 factors (we will use: functional department, age, gender, ethnicity, and team tenure as  
 307 justified later).

308  $P_k^S$  is an index that measures the differential level of performance expected for  
 309 subgroup  $k$  given a particular level of cohesion, which is measured by sociometric indices. In  
 310 this case,  $S_k$  will be calculated as the interpersonal social preferences and rejections stated by  
 311 all members belonging to subgroup  $k$ .

312 What follows are the details on how  $P_k^{SS}$  and  $P_k^S$  are calculated. Once these two  
 313 values are known for each potential subgroup  $k$ , obtaining  $E_k$  will be straightforward with (4).

314 Let us start with  $P_k^{SS}$ , the performance metric coming from the Subgroup Strength  
 315 metric. Conventionally, a subgroup  $k$ 's homogeneity  $h_k$  has been defined as:

$$316 \quad h_k = \frac{\sum_{i<j} \{O_{ij}^d + O_{ij}^a + O_{ij}^g + O_{ij}^e + O_{ij}^t\}}{\frac{n_k(n_k - 1)}{2}} = \frac{\sum_{i<j} O_{ij}}{\frac{n_k(n_k - 1)}{2}} \quad (5)$$

317 Where  $O_{ij}$  is the total overlap between individuals  $i$  and  $j$ , and which is computed as  
 318 the sum of  $O_{ij}^d$ ,  $O_{ij}^a$ ,  $O_{ij}^g$ ,  $O_{ij}^e$  and  $O_{ij}^t$  which, in turn, represent the overlaps between two  
 319 individuals  $i$  and  $j$  on functional department, age, gender, ethnicity and team tenure,  
 320 respectively. The sum in the numerator is restrained to  $i<j$  (but it could have also been  $i>j$   
 321 indistinctly) to avoid the cases where  $i=j$  (individuals' self-overlaps) as well as to prevent the  
 322 symmetrical  $O_{ij}$  values (that is  $O_{ij}=O_{ji}$ ) from being counted twice.

323 Also in the same vein, the factor  $n_k(n_k-1)/2$  in the denominator of (5) corresponds to  
 324 the total number of pairs analyzed (all possible combinations of  $i$  and  $j$ , excluding those cases  
 325 where  $i \geq j$ ).

326 With all this in mind, and according to Gibson and Vermeulen (2003), the different  
 327 overlaps between a group of individuals can be calculated as follows:

328 Functional department overlap:  $O_{ij}^d = 1$  if  $d_i=d_j$ , else 0 (6)

329 Age overlap:  $O_{ij}^a = \frac{\min(a_i, a_j) - 18}{\max(a_i, a_j) - 18}$  (7)

330 Gender overlap:  $O_{ij}^g = 1$  if  $g_i=g_j$ , else 0 (8)

331 Ethnicity overlap:  $O_{ij}^e = 1$  if  $e_i=e_j$ , else 0 (9)

332 Team tenure overlap:  $O_{ij}^t = \frac{\min(t_i, t_j)}{\max(t_i, t_j)}$  (10)

333 Overlap values can vary between [0, 1]. Hence, values of  $h_k$  will vary between [0, 5].  
 334 We are aware that other diversity factors could have also been included in the definition of  $h_k$   
 335 such as for example, language, education, experience. However, in the interest of keeping to  
 336 the model's simplicity and for illustrative purpose in the case study which follows, we  
 337 deemed it reasonable to stick to the diversity factors in the definition of  $h_k$  as proposed by  
 338 Blau (1977) and Allison (1978).

339 And now that the overlaps of all individuals  $O_{ij}$  and the subgroup  $k$ 's homogeneity  
 340 value  $h_k$  have been detailed, the Subgroup Strength of a subgroup  $k$  ( $SS_k$ ) is defined as the  
 341 population standard deviation of the  $O_{ij}$  values from all  $n_k$  members belonging to subgroup  $k$ ,  
 342 that is:

343 
$$SS_k = \sqrt{\frac{\sum_{i < j} (O_{ij} - h_k)^2}{\frac{n_k(n_k - 1)}{2}}} = Std. Dev. O_{ij}$$
 (11)



344 As defined,  $SS_k$  will vary from 0 to 1.25 (since  $h_k$  domain was restricted to  $[0,5]$ ).  
 345 Additionally, Gibson and Vermeulen (2003) proved that team diversity (represented by  
 346 means of  $SS_k$ ) and group performance were quadratically related (inverted U-shape)  
 347 approximately as described in Figure 1a.

348 **<Insert Figure 1 here>**

349 Also, a recent study by Chen et al. (2017) suggested that this quadratic expression is  
 350 quasi-symmetrical and that the value of  $\delta$  seems generally close to 10% on average.  
 351 Therefore, the subgroup  $k$ 's performance  $P_k^{SS}$  can be calculated from the subgroup strength  
 352  $SS_k$  value as:

$$353 \quad P_k^{SS} = -2.56 \delta SS_k^2 + 3.2 \delta SS_k + 1 - \delta \quad \text{with } \delta \approx 0.10 \quad (12)$$

354 Expression (12) can vary between  $[1-\delta, 1]$  and is obtained from a quadratic  
 355 polynomial which is forced to cross the points:  $(0, 1-\delta)$ ,  $(1.25/2, 1)$  and  $(1.25, 1-\delta)$ .

356 On the other hand, the subgroup  $k$ 's cohesion-related performance index  $P_k^S$  is  
 357 calculated from the subgroup  $k$  members' sociometric preferences  $s_{ij}$  towards each other (i.e.  
 358 preferences and rejections to work with a particular individual). These preferences and  
 359 rejections do not have to be symmetrical (that is,  $S_{ij} \neq S_{ji}$  or  $S_{ij} = S_{ji}$ ). Hence, we define a  
 360 subgroup  $k$ 's cohesion  $S_k$  as:

$$361 \quad S_k = \frac{\sum_{i \neq j} s_{ij}}{n_k (n_k - 1)} \quad (13)$$

362 Similarly, the term  $n_k (n_k - 1)$  corresponds to the total number of pairs analyzed  
 363 excluding the choices of individuals with themselves. So, as  $s_{ij}$  can be equal to  $-1$  (meaning  $i$   
 364 dislikes  $j$ ),  $0$  ( $i$  is neutral or have not met  $j$ ), or  $+1$  ( $i$  likes  $j$ ),  $S_k$  actually represents how well

365 (or badly) all subgroup  $k$ 's members get along with each other on average. Analogously,  $S_k$   
366 can take on values within the range  $[-1, 1]$ .

367 Finally, previous researchers' results suggest that the average cohesive group seems  
368 to perform around 18% better than average (non-cohesive or non-uncohesive) groups (Evans  
369 and Dion 2012). For the purpose of this paper, this performance differential will be called  $\varphi$  .  
370 However, it is worth pointing out that in those previous pieces of research it is not always  
371 clear how group cohesion is measured or quantified. Also, there is a total absence of studies  
372 clarifying whether the cohesion-performance relationship is linear or if it indeed follows a  
373 different pattern. In light of this, it seems prudent to take the simplest alternative and assume  
374 that group cohesion (represented now by  $S_k$ ) and performance ( $P_k^S$ ) will just be linearly  
375 related as represented in Figure 1b. Hence:

$$376 \quad P_k^S = 1 + \varphi S_k \quad \text{with } \varphi \approx 0.18 \quad (14)$$

377 After defining expression (13), all variables involved have been presented and related  
378 to each other. We are now able to calculate the global group efficiency (E) from the different  
379 simultaneous subgroups' efficiencies  $E_k$  and their respective weights  $w_k$ . This is summarized  
380 at the bottom of Figure 1. Hence, from now on, every possible subgroups' allocation can be  
381 measured in relative performance terms and each feasible complete group allocation can be  
382 compared against each other. The following is an example to illustrate how we can apply the  
383 model based on a fictitious case study which reflects as much as possible, a real project  
384 environment.

385



410

**<Insert Figure 4 here>**

411

What remains is calculating, for any potential and feasible subgroup  $k$ , its

412

homogeneity  $h_k$  (with expression (5)), subgroup strength  $SS_k$  (with expression (11)), and

413

cohesion  $S_k$  (with expression (13)) values. Then, with  $SS_k$  and  $S_k$ , known, calculating the

414

homogeneity-related  $P_k^{SS}$  performance metric (with expression (12)) and the cohesion-related

415

$P_k^S$  performance metric (with expression (14)) can be performed. Next, with  $P_k^{SS}$  and  $P_k^S$

416

known, we can calculate  $E_k$  (by means of expression (4)). Once the values of  $E_k$  are all known

417

for all the simultaneous subgroups (three in our example, as there are three projects), and by

418

knowing the weight of each subgroup  $w_k$  (with expressions (2) or (3), and as detailed at the

419

bottom of Figure 3), it is possible to obtain the global efficiency  $E$  of that group configuration

420

by means of expression (1). This series of calculations are represented vertically from top to

421

bottom in the lower half of Figure 5.

422

**<Insert Figure 5 here>**

423

In Figure 5, a random solution directly allocating the individuals available to meet the

424

project staffing requirement, but without any further (homogeneity, nor cohesion)

425

considerations, is presented. At the top of Figure 5, one can find the allocation of each

426

individual to each subgroup/project. The column to the right sums each individual's

427

allocations and verifies that no individual is allocated more than 100%, that is, to more than

428

one project. These are necessary but not sufficient problem constraints which need to be met

429

to qualify any allocation as feasible.

430

However, every time there is any change (for instance a member is allocated to a

431

different project/subgroup), all values need to be recalculated. Therefore, the only way of

432

finding good solutions is by iterating these calculations multiple times while testing as many

433 feasible solutions as possible. Hence, for the model to be practically useful, this task must be  
434 automated by an optimization algorithm.

435 The global efficiency of the random solution depicted in Figure 5 corresponds to  
436 0.981. By definition, the value of the global efficiency  $E$  will vary between  $[1-\delta-\phi, 1+\phi]$ .  
437 Hence, the closer the value of  $E$  to  $1+\phi$ , the higher the expected groups' performance. Now,  
438 in looking for that optimum solution, there is one last point to be discussed.

439 As discussed earlier, the proposed model falls within a particular case of the MTFP.  
440 The MTFP is NP-hard, and that opens the door to the use of metaheuristics when looking for  
441 (near) optimum solutions. Among common metaheuristics, Genetic Algorithms (GA) are one  
442 of the quicker and simpler, but also effective, options. On top of that, they are usually  
443 available within commercial solvers included by default in spreadsheet software like  
444 Microsoft Excel ®. Moreover, GA have been in use in resource-constrained allocation for a  
445 long time and they have been considered as one of the most efficient metaheuristics when  
446 dealing with the MTFP in recent studies (e.g. Ahmadian Fard Fini et al. 2016; Awal and  
447 Bharadwaj 2014). For all these reasons, we will use GA for solving multiple instances of our  
448 case study and find a quick and good (despite maybe less than optimal) solution.

449 Therefore, once the model is implemented in a spreadsheet where every time a new  
450 allocation is proposed all the subgroup calculations can be automatically and instantly  
451 updated, the GA can start looking for new solutions. The problem constraints are the ones  
452 specified in grey cells (one per individual's maximum allocation time plus one per project  
453 staffing requirements check). The objective function corresponds to  $E$ , which on this  
454 occasion is to be maximized. On resorting to Excel Solver, the best solution was found in  
455 seconds by our GA as shown in Figure 6.

456 <Insert Figure 6 here>

457           The solution shown in Figure 6 (with  $E=1.087$ ) is comparatively much more efficient  
458 than the one from Figure 5 (with  $E=0.981$ ). If the same GA would have been aimed at  
459 minimizing  $E$ , the worst solution found (not included) would have had an  $E=0.911$ . Within  
460 the  $[1-\delta-\varphi, 1+\varphi]$  interval, the three solutions correspond to the following percentiles: 56.7%  
461 (the random solution), 79.8% (the best solution) and 41.5% (the worst solution). As can be  
462 seen, based on the results, there seems to be valid reasons to try to optimize the model  
463 outputs with the help of an optimization algorithm. Mostly, when doing this manually, would  
464 have been an unsurmountable task.

465

#### 466       **Validation**

467           In the previous section it was shown how the model can be implemented to fulfill its  
468 most common purpose: finding the optimum (or near optimum) allocation of a set of  
469 available human resources into a series of projects, each with not necessarily equal staffing  
470 requirements. However, before accepting that the model outputs constitute a fair description  
471 of reality, it is necessary to verify whether its parameters actually influence different levels of  
472 team performance. Particularly, the most relevant model parameters are the ones proposed in  
473 equations (12) and (14), that is, the group diversity-related performance ( $P_k^{SS}$ ) and the social  
474 cohesion-related performance ( $P_k^S$ ). Hence, if higher values of these two parameters exhibit  
475 correlation with higher values of team performance, the model will be of some value.  
476 Conversely, if there is no such correlation, the model, at least as currently formulated, would  
477 render useless.

478           With this purpose in mind, a first exploratory and validation study was conducted  
479 comprising the academic performance of fifteen MSc Civil Engineering students at the  
480 Universidad de Talca (Chile). This group of students were enrolled in a module named

481 'Projects' which is a transversal integration module and its purpose is to determine how well  
482 students can apply knowledge and understanding from previous related modules. The module  
483 was led by one of the authors in the second semester of 2017. It required that fifteen students  
484 submitted three assignments (projects) each. The three projects which will be named Project  
485 1, Project 2 and Project 3, had progressive submission dates every two months. Students  
486 worked in groups of five to deliver these three projects. After each project was completed, the  
487 groups were reshuffled so that most students had to work with different team mates in the  
488 next project.

489 In short, for the first assignment (Project 1), there were three groups with 5 students  
490 (named here as groups A, B and C) each submitting a different project. The same happened  
491 for Project 2 and 3, but with groups whose member composition was different from Project 1.  
492 Each of these three projects was assessed and given a mark between 0 and 100. In total, there  
493 were 9 different marks: one per assignment (Project 1, 2 and 3) and group (A, B and C).  
494 However, each student only received three marks (one from each Project) whose average  
495 resulted in the module's final mark for him/her.

496 The demographic attributes of the fifteen students can be found in Figure 7. By  
497 columns, the five individuals' attributes had a close equivalence with the five attributes  
498 described in our model: background (akin to functional department), age, gender, ethnicity  
499 and work experience (akin to team tenure). However, as expected from a group of students,  
500 the sample was also relatively more homogeneous than other real-life projects (most  
501 individuals had similar ages, similar experience, and a less varied set of  
502 backgrounds/degrees).

503 **<Insert Figure 7 here>**

504 The group demographic attributes were directly retrieved from their registrations  
505 information. This MSc programme required a minimum work experience of 2 years as an  
506 admission criterion. This is the reason why most students exhibit similar years of experience,  
507 but also similar ages. Additionally, most were local students, which meant most of them were  
508 South American.

509 The Sociometric individuals' preferences matrix was populated using the registration  
510 of the same students for a series of lab sessions which ran in parallel to this module.  
511 Particularly, before allocating the fifteen students to those lab sessions, they were asked with  
512 whom they would like to carry out the lab sessions and whom they would prefer to avoid.  
513 Although it was not compulsory for the instructors to implement those preferences, those  
514 registers proved useful later for allocating the students to the Project groups, and also to  
515 populate the sociometric matrix.

516 Finally, the last three columns in Figure 7 correspond to the three marks that each  
517 student was awarded at the end of the module.

518 The information from Figure 7 is enough to develop a first, simple, and representative  
519 correlation analysis between the model parameters and team performance. For this purpose,  
520 the groups' homogeneity, subgroup strength and social cohesion values were calculated for  
521 the nine five-student groups that submitted the three projects. With these values, calculating  
522 the group diversity- and cohesion-related performances was straightforward. These  
523 calculations are all presented, along with the students' allocations, in Figure 8.

524 **<Insert Figure 8 here>**

525 Values highlighted in blue, red and green correspond to Projects 1, 2 and 3, which  
526 were also identified in the last three columns of Figure 7. For reference purposes, the project  
527 marks were also shown at the bottom row of Figure 8.





552 be fairly representative and is correctly indicating that certain heterogeneity-related and  
553 cohesion-related group attributes can ultimately lead to higher (or lower) team performances.

554         Of course, the conclusions of this validation case study have to be taken with some  
555 caution too. The analysis is based on an academic environment, rather than a real project.  
556 Real life projects tend to consist of a more diverse group of professionals (higher dispersion  
557 of the demographic attributes) with generally many more variables which may be difficult  
558 (but necessary) to control. Notwithstanding this, we acknowledge further validation using  
559 real projects is needed in order to improve the validity of the model. However, resorting to an  
560 academic environment also has numerous advantages. First, the outcome of ‘project’  
561 performance can be known (under some simplifying assumptions) as all assignments are  
562 graded and awarded a mark. And second, these project cycles are usually faster which also  
563 allows data retrieval to be generated faster than in real-life projects.

564         Other limitation of our validation case study is the reduced number of points (only  
565 nine) and the reduced variation of some of the performance measurements. In connection  
566 with the latter, the cohesion-related performance values of Projects 2 and 3 are very close to  
567 each other, obscuring the type of relationship that more dispersed values could have shown.  
568 Also, although the rest of the cases show clearer trends, it is necessary to point out that these  
569 might not be necessarily linear. This, despite us resorting to three points, two of them still  
570 remain too close in the cohesion-related performance graph to infer properly potentially non-  
571 liner trends.

572         Finally, it is clear from Figure 9 that Project 1 seemed to be more challenging to the  
573 students as they all got lower marks (probably because it was the first assignment), whereas  
574 the other two seemed easier (they received higher marks). Similarly, for future validation  
575 studies, it will be advisable to gather individual marks from each student (by means of  
576 individual exams, for example). Only with this additional piece of information, will it be

577 possible to better compare different levels of group members' performance (as groups made  
578 up of bright people usually perform better than ones with mediocre students).

579

## 580 **Discussion**

581 In this paper, a new model for allocating human resources that considers team  
582 functional requirements, group heterogeneity and social cohesion has been proposed and  
583 validated. In formulating the model, a few simplifications and constraints were assumed.  
584 These will be now be reviewed and discussed in detail.

585 First of all, as stated earlier, this model has necessarily oversimplified the nature of  
586 real life work collaboration issues. Real life team work is complex and dynamic. Certainly, it  
587 cannot be reduced to two variables –group diversity and social cohesion– without neglecting  
588 aspects that make from group collaboration something rich and distinct from other  
589 engineering and technical challenges. In real life, group members' exhibit behaviors and  
590 possess attributes that have not been included in this model (e.g. how introvert/extrovert  
591 group members are; their dedication, devotion, preferences or just personal or professional  
592 interests or goals; their soft skills or motivation to work in groups; the asymmetrical personal  
593 relationships as a consequence of the lines of command, etc.). However, the intention here  
594 was not to include an exhaustive list of group attributes but to present a simple and self-  
595 contained model. And despite all the necessary simplifications, the model still seems to be  
596 robust, at least, based on the preliminary validation results shown here. The inclusion of  
597 further variables will be something that, no doubt, will be considered in future versions of the  
598 model when it is applied in real project settings.

599 Secondly, one might raise the question that the way group heterogeneity and cohesion  
600 have been defined in this paper might lead to some collinearity or, at least, covariance

601 between the two constructs. This is because there is a possibility that both may be capturing  
602 some common aspects of a group configuration. Our model, however, has instrumented both  
603 constructs in a multiplicative way, that is,  $P_k^{SS} \times P_k^S$ , not additive, because they do not  
604 substitute for one another when contributing to subgroup performance  $E_k$ . All the same, we  
605 agree that this might be an over simplification, but existing research so far does not seem to  
606 have reached an agreement on whether this is an untenable assumption.

607 For example, Festinger et al. (1950) in an early attempt found contradicting results in  
608 two experiments analyzing the group heterogeneity-cohesion relationship. Much more  
609 recently, Dion (2000), on performing analysis in two houses of war veterans found again  
610 inconclusive findings indicating that in one house cohesion was related with homogeneity,  
611 whereas in the other it was not. And even more recently, in a study conducted by Chiochio  
612 and Essiembre (2009), it was shown that a group's homogeneity or heterogeneity does not  
613 appear to affect the social cohesion-behavioral performance correlations in either academic or  
614 organizational settings. In line with this, probably the most enlightening stance has been the  
615 one taken by Sturgis et al. (2014), who claimed that the relationships between the different  
616 subcomponents of group heterogeneity and cohesion might be very different from each other,  
617 even cancelling out each other's effects. They also emphasized that more research is  
618 necessary to validate this.

619 However, and fortunately, because our model only tries to relatively (not absolutely)  
620 generate the most desirable subgroups allocations from the same pool of human resources,  
621 the effect of potential (if existing) collinearities between subcomponents of heterogeneity and  
622 cohesion will not be that critical so as to invalidate the model. This, as despite correlation  
623 between both variables might cause some scale distortion, the relative rank (order) of  
624 solutions should not have altered much.

625           Additionally, other simplifications have been assumed along the model formulation.  
626   Probably the two most relevant have been limiting the allocations of individuals to be in full-  
627   time working arrangement and not part-time. Also, the relationship between group cohesion  
628   and group performance has been assumed to be linear.

629           The first simplification is relatively easy to address but it would complicate the  
630   mathematical expressions to a point where they are no longer that intuitive. In this paper, we  
631   have tried to encourage understanding of the model's utility and to avoid distractions by  
632   complicating it too much. However, allowing part-time allocations might make finding better  
633   solutions somewhat easier for an optimization algorithm. This, as the objective functions of  
634   many OR models are generally easier to optimize when the decision variables are closer to  
635   being continuous (Gutiérrez et al., 2016).

636           Finally, the second simplification cannot be satisfactorily addressed until there is  
637   more research to determine the nature of the cohesion-performance relationship. This might  
638   be a critical aspect for further model development. It may indeed lead to some adjustments in  
639   some of the equations (probably in expression (4) and surely in expression (14)), but for now  
640   there is no point in us speculating how it might impact the model formulation, or indeed if  
641   there is such an impact at all.

642

## 643           **Conclusions**

644           A model that allocates human resources to multiple projects with specific staffing  
645   requirements while also considering group homogeneity and cohesion has been proposed.  
646   This model constitutes a powerful and practical tool for any project manager who needs to  
647   efficiently allocate human resources and who wants to maximize the expected productivity of  
648   his/her group members. The mathematical expressions are, in general, quite straightforward

649 and can be easily implemented by means of a spreadsheet. The optimization algorithm for  
650 finding near-optimal solutions can also be implemented with the aid of a very simple  
651 commercial solver like Excel Solver (currently a free, despite capped, version of Frontline  
652 Solvers®).

653 Human resources are a key component of project success, but there is a lack of  
654 practical, quantitative tools that allow project managers to efficiently allocate these resources  
655 and build high performing teams. There are many reasons that can keep a team from  
656 functioning effectively. In this paper, two factors that are found to strongly and consistently  
657 influence group performance – group homogeneity and group cohesion – have been  
658 incorporated within the model. This model allows the measuring and comparing of any set of  
659 feasible subgroup allocations to several projects simultaneously.

660 Namely, group homogeneity has been defined by the subgroup strength metric and the  
661 sum of overlaps between subgroup members on five different demographic sub-factors  
662 (functional department, age, sex, ethnicity and team tenure). Group cohesion has been  
663 defined as the degree of acceptance (or rejection) that all members have with each other. The  
664 information on the five sub-factors in the group cohesion construct is generally very easy to  
665 obtain from the group members' professional profiles. In terms of the degree of  
666 acceptance/rejection that each group member has toward the rest of their group members,  
667 these can generally be known by using sociometric questionnaires. The latter, despite its  
668 limitations, have also been proven in previous research to be quite representative and  
669 relatively easy to use and update. Basically, these questionnaires require asking all group  
670 members who have finished a project: "Who would you like to keep working with?" and  
671 "Who would you prefer not to work with? From the group members' answers it is possible to  
672 populate (and keep updating) a sociometric matrix that is eventually useful for measuring  
673 how cohesive each potential subgroup is or can be.

674 Furthermore, previous research has proven that group homogeneity can  
675 reduce/increase group performance by up to 10% on average. Similarly, group cohesion is  
676 responsible for average increases (or decreases) of group performance by up to 18%. Both  
677 figures have been included in the proposed model and allow the objective measuring of the  
678 relative group performance differences between multiple feasible subgroups. Feasible  
679 subgroups are those who fulfill the minimum project staffing requirements stated by some  
680 simultaneous projects.

681 With all this, the proposed mathematical model has been detailed concerning all its  
682 components and variable relationships. A fictitious case study involving twenty workers who  
683 are allocated to three projects have been proposed and solved by means of a simple Genetic  
684 Algorithm. Finally, a validation case study based on an academic setting has also been  
685 included which involved fifteen MSc students who were allocated to three groups and were  
686 required to complete three sequential projects.

687 The proposed model is a simple and yet powerful way of addressing the  
688 commonplace challenges of a typical project manager in efficiently allocating human  
689 resources in projects. Despite some intentional simplifications, the model shows promise in  
690 helping project managers to make more objective and efficient decisions about their human  
691 resource allocations. However, more validating studies will be required in the future to test  
692 the actual utility of the model in real project contexts.

693 Although validation with real projects is necessary, this will also increase the  
694 complexity of the model's application due to the number of variables to be considered, as  
695 well as generally bigger team sizes. Indeed, this wider range of variables will have some  
696 human resource implications in terms of the structure of social cohesion of individuals within  
697 the projects. For example, the interactions amongst members in real projects may be  
698 underpinned by career-related imperatives, and hence, are likely to be more dynamic and

699 nuanced, when compared with students'. In this vein, a potentially fruitful avenue for future  
700 research would be to use real life projects in conjunction with using academic projects as  
701 controlled experiment to enable researchers to study the nature and structure of social  
702 cohesion more precisely.

703

#### 704 **Data availability**

705 All data generated or analyzed during the study are included in the submitted article  
706 or supplemental materials files.

707

#### 708 **References**

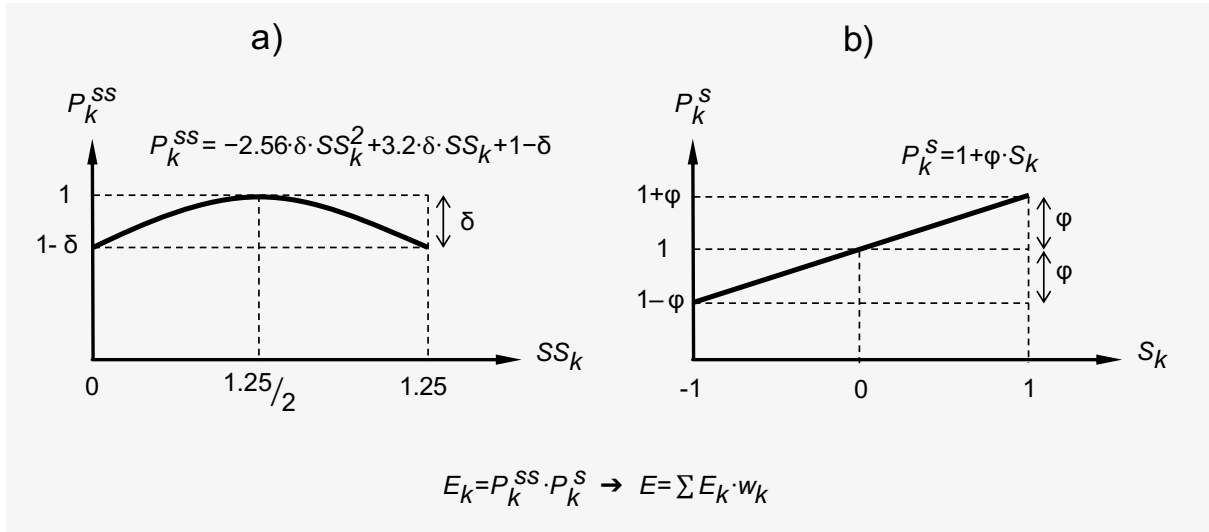
- 709 Agrawal, R., Golshan, B., and Terzi, E. (2014). "Grouping students in educational settings."  
710 *Proceedings of the 20th ACM SIGKDD international conference on Knowledge*  
711 *discovery and data mining - KDD '14*, ACM Press, New York, New York, USA, 1017–  
712 1026.
- 713 Ahmadian Fard Fini, A., Akbarnezhad, A., Rashidi, T. H., and Waller, S. T. (2017). "Job  
714 Assignment Based on Brain Demands and Human Resource Strategies." *Journal of*  
715 *Construction Engineering and Management*, 143(5), 4016123.
- 716 Ahmadian Fard Fini, A., Rashidi, T. H., Akbarnezhad, A., and Travis Waller, S. (2016).  
717 "Incorporating Multiskilling and Learning in the Optimization of Crew Composition."  
718 *Journal of Construction Engineering and Management*, 142(5), 4015106.
- 719 Al-Bayati, A. J., Abudayyeh, O., Fredericks, T., and Butt, S. E. (2017). "Managing Cultural  
720 Diversity at U.S. Construction Sites: Hispanic Workers' Perspectives." *Journal of*  
721 *Construction Engineering and Management*, 143(9), 4017064.
- 722 Allison, P. D. (1978). "Measure of inequality." *American Sociological Review*, 43, 865–880.
- 723 Anvuur, A. M., and Kumaraswamy, M. M. (2016). "Effects of Teamwork Climate on  
724 Cooperation in Crossfunctional Temporary Multi-Organization Workgroups." *Journal*  
725 *of Construction Engineering and Management*, 142(1), 4015054.
- 726 Awal, G. K., and Bharadwaj, K. K. (2014). "Team formation in social networks based on  
727 collective intelligence – an evolutionary approach." *Applied Intelligence*, Kluwer  
728 Academic Publishers, 41(2), 627–648.
- 729 Ballesteros-Pérez, P., González-Cruz, M. C., and Fernández-Diego, M. (2012). "Human  
730 resource allocation management in multiple projects using sociometric techniques."  
731 *International Journal of Project Management*, 30(8).
- 732 Beal, D. J., Cohen, R. R., Burke, M. J., and McLendon, C. L. (2003). "Cohesion and



- 733 performance in groups: a meta-analytic clarification of construct relations.” *The Journal*  
734 *of applied psychology*, 88(6), 989–1004.
- 735 Blau, P. M. (1977). *Inequality and Heterogeneity*. Free press, New York.
- 736 Castaño, N., Watts, T., and Tekleab, A. G. (2013). “A reexamination of the cohesion-  
737 performance relationship meta-analyses: A comprehensive approach.” *Group Dynamics: Theory, Research and Practice*, 17(4), 207–231.
- 739 Chang, A., and Bordia, P. (2001). “A Multidimensional approach to the Group Cohesion –  
740 Group Performance relationship.” *Small Group Research*, 32(4), 379–405.
- 741 Chen, S., Wang, D., Zhou, Y., Chen, Z., and Wu, D. (2017a). “When too little or too much  
742 hurts: Evidence for a curvilinear relationship between team faultlines and performance.”  
743 *Asia Pacific Journal of Management*, *Asia Pacific Journal of Management*, 34(4), 931–  
744 950.
- 745 Chen, Y., McCabe, B., and Hyatt, D. (2017b). “Relationship between Individual Resilience,  
746 Interpersonal Conflicts at Work, and Safety Outcomes of Construction Workers.”  
747 *Journal of Construction Engineering and Management*, 143(8), 4017042.
- 748 Chiochio, F., and Essiembre, H. (2009). *Cohesion and Performance: A Meta-Analytic*  
749 *Review of Disparities Between Project Teams, Production Teams, and Service Teams.*  
750 *Small Group Research*.
- 751 Dion, K. L. (2000). “Group cohesion: From ‘Field of Forces’ to Multidimensional  
752 Construct.” *Group Dynamics: Theory, Research, and Practice*, 4(1), 7–26.
- 753 Dorn, C., and Dustdar, S. (2010). *On the Move to Meaningful Internet Systems: OTM 2010.*  
754 *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial*  
755 *Intelligence and Lecture Notes in Bioinformatics)*, Lecture Notes in Computer Science,  
756 (R. Meersman, T. Dillon, and P. Herrero, eds.), Springer Berlin Heidelberg, Berlin,  
757 Heidelberg.
- 758 Earley, P. C., and Gibson, C. B. (2002). *Multinational Work Teams: A New Perspective.*  
759 Mahwah, N.J.: Lawrence Erlbaum Associates.
- 760 Evans, C. R., and Dion, K. L. (2012). “Group Cohesion and Performance: A Meta-Analysis.”  
761 *Small Group Research*, 43(6), 690–701.
- 762 Farhadi, F., Hoseini, E., Hashemi, S., and Hamzeh, A. (2012a). “TeamFinder: A Co-  
763 clustering based Framework for Finding an Effective Team of Experts in Social  
764 Networks.” *2012 IEEE 12th International Conference on Data Mining Workshops*,  
765 IEEE, 107–114.
- 766 Farhadi, F., Sorkhi, M., Hashemi, S., and Hamzeh, A. (2011). “An Effective Expert Team  
767 Formation in Social Networks Based on Skill Grading.” *2011 IEEE 11th International*  
768 *Conference on Data Mining Workshops*, IEEE, 366–372.
- 769 Farhadi, F., Sorkhi, M., Hashemi, S., and Hamzeh, A. (2012b). “An Effective Framework for  
770 Fast Expert Mining in Collaboration Networks: A Group-Oriented and Cost-Based  
771 Method.” *Journal of Computer Science and Technology*, 27(3), 577–590.
- 772 Festinger, L., Schachter, S., and Back, K. (1950). *Social pressures in informal groups; a*  
773 *study of human factors in housing*. Harper, Oxford, England.
- 774 Gajewar, A., and Sarma, A. Das. (2012). “Multi-skill collaborative teams based on densest  
775 subgraphs.” *Proceedings of the 12th SIAM International Conference on Data Mining*,  
776 *SDM 2012*, 165–176.

- 777 Gibson, C., and Vermeulen, F. (2003). “A Healthy Divide : Subgroups as a Stimulus for  
778 Team Learning Behavior.” *Administrative Science Quarterly*, 48(2), 202–239.
- 779 Gutiérrez, J. H., Astudillo, C. A., Ballesteros-Pérez, P., Mora-Melià, D., and Candia-Véjar,  
780 A. (2016a). “The multiple team formation problem using sociometry.” *Computers and  
781 Operations Research*, 75.
- 782 Gutiérrez, J. H., Astudillo, C. A., Ballesteros-Pérez, P., Mora-Melià, D., and Candia-Véjar,  
783 A. (2016b). “The multiple team formation problem using sociometry.” *Computers and  
784 Operations Research*, Elsevier, 75, 150–162.
- 785 Hendriks, M. H. A., Voeten, B., and Kroep, L. (1999). “Human resource allocation in a  
786 multi-project R&D environment: Resource capacity allocation and project portfolio  
787 planning in practice.” 17(3), 181–188.
- 788 Lappas, T., Liu, K., and Terzi, E. (2009). “Finding a team of experts in social networks.”  
789 *Proceedings of the 15th ACM SIGKDD international conference on Knowledge  
790 discovery and data mining - KDD '09*, ACM Press, New York, New York, USA, 467.
- 791 Lau, D. C., and Murnighan, J. K. (2005). “Interactions within Groups and Subgroups: The  
792 Effects of Demographic Faultlines.” *The Academy of Management Journal*, Academy of  
793 Management.
- 794 Li, C.-T., and Shan, M.-K. (2010). “Team Formation for Generalized Tasks in Expertise  
795 Social Networks.” *2010 IEEE Second International Conference on Social Computing*,  
796 IEEE, 9–16.
- 797 Lott, A. J., and Lott, B. E. (1965). “Group cohesiveness as interpersonal attraction: A review  
798 of relationships with antecedent and consequent variables.” *Psychological Bulletin*,  
799 64(4), 259–309.
- 800 Mathieu, J. E., Kukenberger, M. R., D’Innocenzo, L., and Reilly, G. (2015). “Modeling  
801 Reciprocal Team Cohesion – Performance Relationships , as Impacted by Shared  
802 Leadership and Members ’ Competence Modeling Reciprocal Team Cohesion –  
803 Performance Relationships , as Impacted by Shared Leadership and Me.” *Journal of  
804 Applied Psychology*, 100(3), 713–734.
- 805 Meyer, B., and Glenz, A. (2013). “Team Faultline Measures: A Computational Comparison  
806 and a New Approach to Multiple Subgroups.” *Organizational Research Methods*, 16(3),  
807 393–424.
- 808 Moreno, J. L. (1941). “Foundations of Sociometry: An Introduction.” *Sociometry*, 4(1), 15–  
809 35.
- 810 Phua, F. T. T. (2004). “The antecedents of co-operative behaviour among project team  
811 members: an alternative perspective on an old issue.” *Construction Management and  
812 Economics*, Taylor & Francis Group , 22(10), 1033–1045.
- 813 Phua, F. T. T., and Rowlinson, S. (2004). “How important is cooperation to construction  
814 project success? A grounded empirical quantification.” *Engineering, Construction and  
815 Architectural Management*, Emerald Group Publishing Limited, 11(1), 45–54.
- 816 Salo, M. (2006). *The Relation Between Sociometric Choices and Group Cohesion*. Arlington,  
817 Virginia.
- 818 Shi, Z., and Hao, F. (2013). “A strategy of multi-criteria decision-making task ranking in  
819 social-networks.” *The Journal of Supercomputing*, 66(1), 556–571.
- 820 Sorkhi, M., Alvari, H., Hashemi, S., and Hamzeh, A. (2012). “A game-theoretic framework

- 821 to identify top-k teams in social networks.” *KDIR 2012 - Proceedings of the*  
822 *International Conference on Knowledge Discovery and Information Retrieval*, 252–257.
- 823 Sturgis, P., Brunton-Smith, I., Kuha, J., and Jackson, J. (2014). “Ethnic diversity, segregation  
824 and the social cohesion of neighbourhoods in London.” *Ethnic and Racial Studies*,  
825 Taylor & Francis, 37(8), 1286–1309.
- 826 Teixeira, L. O., and Huzita, E. H. M. (2014). *DiSEN-AlocaHR: A Multi-Agent Mechanism for*  
827 *Human Resources Allocation in a Distributed Software Development Environment*  
828 *International Conference*. Advances in Intelligent Systems and Computing, (S. Omatu,  
829 H. Bersini, J. M. Corchado, S. Rodríguez, P. Pawlewski, and E. Bucciarelli, eds.),  
830 Springer International Publishing, Cham.
- 831 Tseng, T.-L. (Bill), Huang, C.-C., Chu, H.-W., and Gung, R. R. (2004). “Novel approach to  
832 multi-functional project team formation.” *International Journal of Project Management*,  
833 22(2), 147–159.
- 834 Xia, N., Zhong, R., Wu, C., Wang, X., and Wang, S. (2017). “Assessment of Stakeholder-  
835 Related Risks in Construction Projects: Integrated Analyses of Risk Attributes and  
836 Stakeholder Influences.” *Journal of Construction Engineering and Management*, 143(8),  
837 4017030.
- 838 Yin, H., Cui, B., and Huang, Y. (2011). *Advanced Data Mining and Applications. Lecture*  
839 *Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence*  
840 *and Lecture Notes in Bioinformatics)*, Lecture Notes in Computer Science, (J. Tang, I.  
841 King, L. Chen, and J. Wang, eds.), Springer Berlin Heidelberg, Berlin, Heidelberg.



**Fig. 1.** Team  $k$ 's relative homogeneity-related performance (a) and relative cohesion-related performance (b) calculations

Group homogeneity-related information							Sociometric preference towards individual j (s <sub>ij</sub> )																				
Indiv. (i)	Prof. level (l <sub>i</sub> )	Funct. dept. (d <sub>i</sub> )	Age (a <sub>i</sub> )	Gender (g <sub>i</sub> )	Ethnicity (e <sub>i</sub> )	Te. ten. (t <sub>i</sub> )	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	Senior	Architecture	60	Male	European	25		0	-1	0	1	1	1	-1	0	1	1	0	0	0	1	0	1	0	0	0	
2	Senior	Architecture	50	Female	European	20	0		0	1	0	0	0	1	1	-1	0	0	1	1	1	1	1	1	1	0	0
3	Senior	Architecture	40	Male	North Amer.	10	0	1		1	1	1	1	0	1	1	1	1	1	-1	0	0	1	0	0	0	
4	Intermediate	Architecture	35	Male	North Amer.	5	0	1	1		0	0	1	1	1	1	0	-1	0	1	1	0	1	0	0	0	
5	Junior	Architecture	25	Female	Asian	5	1	1	0	1		1	0	1	1	0	0	0	0	0	0	1	0	0	0	-1	
6	Senior	Civil eng.	55	Male	Asian	10	0	1	1	1	-1		1	0	1	0	0	0	0	1	0	0	1	0	0	1	
7	Senior	Civil eng.	45	Male	Australia	15	0	0	-1	-1	0	1		1	0	0	0	1	-1	0	-1	1	1	1	0	1	
8	Intermediate	Civil eng.	45	Male	European	15	0	1	-1	0	0	1	1		1	-1	0	0	0	1	1	1	0	1	0		
9	Intermediate	Civil eng.	35	Male	European	2	0	0	1	1	-1	-1	1	0		-1	1	-1	0	1	1	0	0	0	0		
10	Junior	Civil eng.	30	Female	African	1	0	0	1	-1	1	-1	0	0	1		1	0	1	0	1	0	1	0	1	-1	
11	Senior	Mechanical eng.	65	Male	European	40	0	0	1	-1	0	0	-1	1	0	-1		1	0	1	0	0	1	1	-1	0	
12	Senior	Mechanical eng.	60	Male	European	35	0	0	0	0	-1	0	0	1	0	1	0		-1	-1	0	1	0	1	0	-1	
13	Intermediate	Mechanical eng.	35	Male	North Amer.	15	0	1	0	0	1	0	0	1	1	-1	0	0		1	0	1	0	-1	0	1	
14	Junior	Mechanical eng.	35	Female	Australia	5	1	0	-1	1	1	0	-1	0	0	0	1	0	1		-1	0	-1	0	1	1	
15	Junior	Mechanical eng.	25	Male	European	5	1	0	0	0	0	0	0	0	0	1	1	1	1	0		0	-1	0	0	1	
16	Senior	Electrical eng.	50	Male	European	20	-1	0	1	1	0	0	0	1	1	0	0	-1	-1	1	1		0	0	0	1	
17	Senior	Electrical eng.	45	Male	Australia	15	0	0	0	0	0	0	0	0	0	-1	0	1	-1	0	1	0		1	0	0	
18	Senior	Electrical eng.	40	Male	Asian	10	-1	-1	0	1	0	0	0	0	0	-1	0	1	0	1	0	-1	0		1	0	
19	Intermediate	Electrical eng.	35	Male	South Amer.	10	-1	1	0	1	0	1	0	0	1	1	-1	-1	0	0	0	-1	1	1		0	
20	Intermediate	Electrical eng.	30	Female	North Amer.	10	1	0	0	-1	0	1	1	1	0	1	1	1	1	1	-1	1	1	0	-1		

Fig. 2. Group of resources' professional, homogeneity and sociometric information

		<b>Project k minimum staff requirements (<math>p_k</math>)</b>		
<b>Funct. Dept. (<math>d_i</math>)</b>	<b>Prof. level (<math>l_i</math>)</b>	<b>1</b>	<b>2</b>	<b>3</b>
Architecture	Senior	1	0	1
	Intermediate	0	1	0
	Junior	0	0	1
Civil engineering	Senior	0	1	0
	Intermediate	1	0	1
	Junior	0	1	0
Mechanical engineering	Senior	1	0	0
	Intermediate	0	1	0
	Junior	0	0	2
Electrical engineering	Senior	0	0	1
	Intermediate	1	1	1
	Junior	1	0	0

<b>Project k budget (<math>b_k</math>)</b>	\$2,500,000	\$1,500,000	\$1,000,000
<b>Project k weight (<math>w_k</math>)</b>	<b>0.50</b>	<b>0.30</b>	<b>0.20</b>

**Fig. 3.** Minimum project staffing requirements

		Total overlaps ( $O_{ij}=O_{ij}^d+O_{ij}^a+O_{ij}^g+O_{ij}^e+O_{ij}^t$ )																			
Indiv. $i \rightarrow$	$j \rightarrow$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1			3.56	2.92	2.60	1.37	2.28	2.24	3.24	2.48	0.33	3.52	3.71	2.00	0.60	2.37	3.56	2.24	1.92	1.80	0.69
2				2.19	1.78	2.47	1.36	1.59	2.59	1.63	1.43	2.18	2.33	1.28	1.78	1.47	3.00	1.59	1.19	1.03	1.88
3					4.27	1.82	2.59	2.48	2.48	1.97	0.65	1.72	1.81	3.44	1.27	1.82	2.19	2.48	3.00	2.77	2.55
4						2.41	1.96	1.96	1.96	2.40	0.91	1.49	1.55	3.33	2.00	2.41	1.78	1.96	2.27	2.50	2.21
5							1.69	0.59	0.59	0.81	1.78	0.27	0.31	0.75	2.41	2.00	0.47	0.59	1.82	0.91	2.08
6								3.40	3.40	2.66	1.42	2.04	2.17	2.13	0.96	1.69	2.36	2.40	3.59	2.46	1.32
7									4.00	2.76	1.51	1.95	2.07	2.63	1.96	1.59	2.59	4.00	2.48	2.30	1.11
8										3.76	1.51	2.95	3.07	2.63	0.96	2.59	3.59	3.00	2.48	2.30	1.11
9											2.21	2.41	2.46	2.13	1.40	2.81	2.63	1.76	1.97	2.20	0.91
10												0.28	0.31	0.77	1.91	0.78	0.43	0.51	0.65	0.81	2.10
11													4.77	2.74	1.49	3.27	3.18	1.95	1.72	1.61	0.51
12														2.83	1.55	3.31	3.33	2.07	1.81	1.69	0.57
13															2.33	2.75	2.28	2.63	2.44	2.67	2.37
14																2.41	0.78	1.96	1.27	1.50	2.21
15																	2.47	1.59	1.82	1.91	1.08
16																		3.59	3.19	3.03	1.88
17																			3.48	3.30	2.11
18																				3.77	2.55
19																					2.71
20																					

Fig. 4. Total overlaps calculations

Staff  $i$  allocations to each project  $p_k$

Indiv. ( $i$ )	Prof. level ( $l_i$ )	Funct. Dept. ( $d_i$ )	1	2	3	Allocated time
1	Senior	Architecture	1	0	0	1
2	Senior	Architecture	0	0	1	1
3	Senior	Architecture	0	0	0	0
4	Intermediate	Architecture	0	1	0	1
5	Junior	Architecture	0	0	1	1
6	Senior	Civil engineering	0	1	0	1
7	Senior	Civil engineering	0	0	0	0
8	Intermediate	Civil engineering	1	0	0	1
9	Intermediate	Civil engineering	0	0	1	1
10	Junior	Civil engineering	0	1	0	1
11	Senior	Mechanical engineering	1	0	0	1
12	Senior	Mechanical engineering	0	0	0	0
13	Intermediate	Mechanical engineering	0	1	0	1
14	Junior	Mechanical engineering	0	0	1	1
15	Junior	Mechanical engineering	0	0	1	1
16	Senior	Electrical engineering	0	0	1	1
17	Senior	Electrical engineering	0	0	1	1
18	Senior	Electrical engineering	0	1	0	1
19	Intermediate	Electrical engineering	1	0	0	1
20	Intermediate	Electrical engineering	1	0	0	1

Staff requirements check (1/0)

1	1	1
Team $k$ 's size ( $m_k$ )	5	7

Constraints

Team  $k$ 's homogeneity ( $h_k$ )

2.04	1.95	1.89
------	------	------

Team  $k$ 's Subgroup Strength ( $SS_k$ )

1.01	0.97	0.80
------	------	------

Diversity-related perform. ( $P^{SS}_k = -2.56\delta SS_k^2 + 3.25SS_k + 1 - \delta$ )

0.96	0.97	0.99
------	------	------

Team  $k$ 's social cohesion ( $S_k$ )

0.05	-0.05	0.26
------	-------	------

Cohesion-related performance ( $P^S_k = 1 + \phi \cdot S_k$ )

1.01	0.99	1.05
------	------	------

Team  $k$ 's Efficiency ( $E_k = P^{SS}_k \cdot P^S_k$ )

0.97	0.96	1.04
------	------	------

Project  $k$  weight ( $w_k$ )

0.50	0.30	0.20
------	------	------

Global allocation Efficiency ( $E = SE_k \cdot w_k$ ) (maximise) = **0.981**  $[1 - \delta - \phi, 1 + \phi] = [0.72, 1.18]$

Fig. 5. An example of a feasible resources' allocation



Staff *i* allocations to each project  $p_k$

Indiv. ( <i>i</i> )	Prof. level ( $l_i$ )	Funct. Dept. ( $d_i$ )	1	2	3	Allocated time
1	Senior	Architecture	0	0	0	0
2	Senior	Architecture	0	0	1	1
3	Senior	Architecture	1	0	0	1
4	Intermediate	Architecture	0	1	0	1
5	Junior	Architecture	0	0	0	0
6	Senior	Civil engineering	1	0	0	1
7	Senior	Civil engineering	0	1	0	1
8	Intermediate	Civil engineering	0	0	1	1
9	Intermediate	Civil engineering	0	1	0	1
10	Junior	Civil engineering	0	0	0	0
11	Senior	Mechanical engineering	1	0	0	1
12	Senior	Mechanical engineering	0	0	1	1
13	Intermediate	Mechanical engineering	0	1	0	1
14	Junior	Mechanical engineering	0	1	0	1
15	Junior	Mechanical engineering	0	0	1	1
16	Senior	Electrical engineering	0	0	1	1
17	Senior	Electrical engineering	0	0	0	0
18	Senior	Electrical engineering	0	1	0	1
19	Intermediate	Electrical engineering	0	1	0	1
20	Intermediate	Electrical engineering	1	0	0	1

Staff requirements check (1/0)

1	1	1	
Team $k$ 's size ( $m_k$ )	4	7	5

Constraints

Team  $k$ 's homogeneity ( $h_k$ )

1.79	2.30	2.78
------	------	------

Team  $k$ 's Subgroup Strength ( $SS_k$ )

0.72	0.57	0.59
------	------	------

Diversity-related perform. ( $P^{SS}_k = -2.56\delta SS_k^2 + 3.25SS_k + 1 - \delta$ )

1.00	1.00	1.00
------	------	------

Team  $k$ 's social cohesion ( $S_k$ )

0.58	0.33	0.50
------	------	------

Cohesion-related performance ( $P^S_k = 1 + \phi \cdot S_k$ )

1.11	1.06	1.09
------	------	------

Team  $k$ 's Efficiency ( $E_k = P^{SS}_k \cdot P^S_k$ )

1.10	1.06	1.09
------	------	------

Project  $k$  weight ( $w_k$ )

0.50	0.30	0.20
------	------	------

Global allocation Efficiency ( $E = SE_k \cdot w_k$ ) (maximise) = **1.087**  $[1 - \delta - \phi, 1 + \phi] = [0.72, 1.18]$

Fig. 6. Best feasible solution (found) of the case study

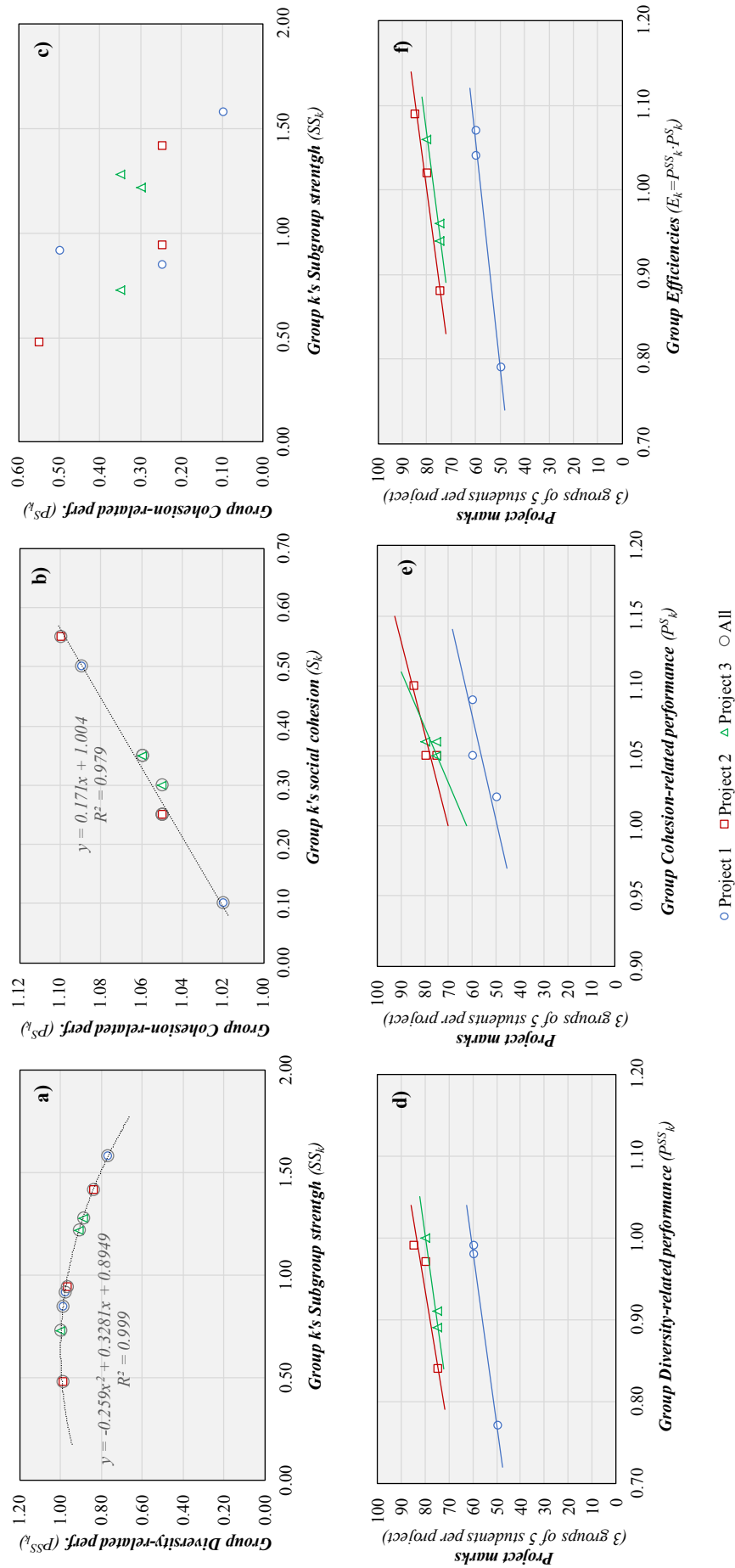
Stud. (i)	Group homogeneity-related information					Sociometric preference towards individual J (s <sub>ij</sub> )															Project group marks*		
	Background (d <sub>i</sub> )	Age (a <sub>i</sub> )	Gender (g <sub>i</sub> )	Ethnicity (e <sub>i</sub> )	Work exp. (t <sub>i</sub> )	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Project 1	Project 2	Project 3
1	Construction	24	Male	South Amer.	2	1	0	1	1	0	1	0	1	0	0	-1	0	0	0	0	60	80	75
2	Construction	24	Female	South Amer.	2	1	0	1	1	0	0	1	0	0	-1	0	0	0	1	50	75	75	
3	Hydraulics	25	Male	South Amer.	3	0	0	1	1	0	0	0	0	0	-1	0	0	0	1	60	85	80	
4	Hydraulics	25	Male	South Amer.	2	1	1	1	1	0	1	0	0	0	0	0	0	0	1	60	85	75	
5	Construction	24	Female	South Amer.	2	1	-1	1	1	0	1	0	0	0	0	0	0	0	0	50	80	80	
6	Construction	24	Male	South Amer.	2	1	0	1	1	1	1	0	0	0	1	0	0	0	0	60	75	75	
7	Hydraulics	25	Female	European	2	0	0	0	0	0	1	1	1	1	0	0	1	0	0	60	80	75	
8	Construction	24	Female	South Amer.	2	1	0	1	1	0	1	1	1	1	0	1	0	0	1	50	75	75	
9	Transport	24	Male	South Amer.	2	0	0	0	0	0	1	1	1	1	1	0	1	0	0	60	85	80	
10	Hydraulics	24	Male	South Amer.	2	0	0	1	0	0	1	1	1	1	1	0	1	0	1	60	85	75	
11	Transport	30	Male	South Amer.	5	0	0	0	0	1	0	0	1	1	1	1	-1	0	1	50	80	80	
12	Hydraulics	24	Female	European	2	0	0	0	0	1	1	1	0	0	-1	0	0	0	0	60	75	75	
13	Construction	24	Male	Asian	2	0	0	0	0	0	1	0	1	1	0	0	0	0	1	60	80	75	
14	Transport	57	Male	South Amer.	30	0	0	0	0	0	0	0	0	0	1	0	0	0	0	50	75	75	
15	Hydraulics	24	Male	South Amer.	2	0	1	1	1	0	0	0	1	0	1	1	1	1	0	60	85	80	

\* Note: all students were allocated to groups of five students to develop three projects. On completing each project, student groups were reshuffled so that groups composition changed for the next assignment.

Fig. 7. Group demographics and marks awarded in the MSc Civil Engineering 'Projects' integration module

Student (i) ↓ & Group (k) →	Project 1			Project 2			Project 3		
	A	B	C	A	B	C	A	B	C
1	1	0	0	0	0	1	0	1	0
2	0	1	0	1	0	0	0	0	1
3	0	0	1	0	1	0	1	0	0
4	1	0	0	0	1	0	0	0	1
5	0	1	0	0	0	1	1	0	0
6	0	0	1	1	0	0	0	1	0
7	1	0	0	0	0	1	0	1	0
8	0	1	0	1	0	0	0	0	1
9	0	0	1	0	1	0	1	0	0
10	1	0	0	0	1	0	0	0	1
11	0	1	0	0	0	1	1	0	0
12	0	0	1	1	0	0	0	1	0
13	1	0	0	0	0	1	0	1	0
14	0	1	0	1	0	0	0	0	1
15	0	0	1	0	1	0	1	0	0
<b>Group k's size (m<sub>k</sub>)</b>	5	5	5	5	5	5	5	5	5
<b>Group k's homogeneity (h<sub>k</sub>)</b>	3.21	2.78	3.31	2.59	4.38	2.53	3.25	2.94	2.85
<b>Group k's Subgroup Strength (SS<sub>k</sub>)</b>	0.92	1.58	0.84	1.41	0.48	0.94	0.73	1.28	1.22
<b>Diversity-rel. perform. (P<sup>SS</sup><sub>k</sub>=-2.56δSS<sub>k</sub><sup>2</sup>+3.2δSS<sub>k</sub>+1-δ)</b>	0.98	0.77	0.99	0.84	0.99	0.97	1.00	0.89	0.91
<b>Group k's social cohesion (S<sub>k</sub>)</b>	0.50	0.10	0.25	0.25	0.55	0.25	0.35	0.35	0.30
<b>Cohesion-related performance (P<sup>S</sup><sub>k</sub>=1+φ·S<sub>k</sub>)</b>	1.09	1.02	1.05	1.05	1.10	1.05	1.06	1.06	1.05
<b>Group k's Efficiency (E<sub>k</sub>=P<sup>SS</sup><sub>k</sub>·P<sup>S</sup><sub>k</sub>)</b>	1.07	0.79	1.04	0.88	1.09	1.02	1.06	0.94	0.96
<b>Project marks (by groups) (from Fig. 7)</b>	60	50	60	75	85	80	80	75	75

**Fig. 8.** Group diversity-related performances, Cohesion-related performances and Group Efficiencies calculations.



**Fig. 9.** Different combinations of Group Diversity-related performance, Group Cohesion-related performance and Group Efficiencies with Project group marks regression plots.

## Supplemental Online Material

This addendum presents the tables with the detailed calculations of the case study members' overlaps concerning functional department, age, gender, ethnicity and team tenure (in the same order).

The complete case study along with all calculations can be downloaded in an excel file here: <http://bit.ly/2BXCqQj> . Additionally, the validation case study can be downloaded here: <http://bit.ly/2yOa9KH>

Functional department overlaps ( $O^d_{ij}$ )																				
Indiv. $i$ & $j \rightarrow$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1		1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	1		1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	1		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	1	1	1		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0		1	1	1	1	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	1		1	1	1	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	1	1		1	1	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	1	1	1		1	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	1	1	1	1		0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0		1	1	1	1	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	1		1	1	1	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	1	1		1	1	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	1	1	1		1	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	1	1	1	1		0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		1	1	1	1
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1		1	1	1
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1		1	1
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1		1
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	

Age overlaps ( $O^a_{ij}$ )																				
Indiv. $i$ & $j \rightarrow$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		0.76	0.52	0.40	0.17	0.88	0.64	0.64	0.40	0.29	0.89	1.00	0.40	0.40	0.17	0.76	0.64	0.52	0.40	0.29
2	0.76		0.69	0.53	0.22	0.86	0.84	0.84	0.53	0.38	0.68	0.76	0.53	0.53	0.22	1.00	0.84	0.69	0.53	0.38
3	0.52	0.69		0.77	0.32	0.59	0.81	0.81	0.77	0.55	0.47	0.52	0.77	0.77	0.32	0.69	0.81	1.00	0.77	0.55
4	0.40	0.53	0.77		0.41	0.46	0.63	0.63	1.00	0.71	0.36	0.40	1.00	1.00	0.41	0.53	0.63	0.77	1.00	0.71
5	0.17	0.22	0.32	0.41		0.19	0.26	0.26	0.41	0.58	0.15	0.17	0.41	0.41	1.00	0.22	0.26	0.32	0.41	0.58
6	0.88	0.86	0.59	0.46	0.19		0.73	0.73	0.46	0.32	0.79	0.88	0.46	0.46	0.19	0.86	0.73	0.59	0.46	0.32
7	0.64	0.84	0.81	0.63	0.26	0.73		1.00	0.63	0.44	0.57	0.64	0.63	0.63	0.26	0.84	1.00	0.81	0.63	0.44
8	0.64	0.84	0.81	0.63	0.26	0.73	1.00		0.63	0.44	0.57	0.64	0.63	0.63	0.26	0.84	1.00	0.81	0.63	0.44
9	0.40	0.53	0.77	1.00	0.41	0.46	0.63	0.63		0.71	0.36	0.40	1.00	1.00	0.41	0.53	0.63	0.77	1.00	0.71
10	0.29	0.38	0.55	0.71	0.58	0.32	0.44	0.44	0.71		0.26	0.29	0.71	0.71	0.58	0.38	0.44	0.55	0.71	1.00
11	0.89	0.68	0.47	0.36	0.15	0.79	0.57	0.57	0.36	0.26		0.89	0.36	0.36	0.15	0.68	0.57	0.47	0.36	0.26
12	1.00	0.76	0.52	0.40	0.17	0.88	0.64	0.64	0.40	0.29	0.89		0.40	0.40	0.17	0.76	0.64	0.52	0.40	0.29
13	0.40	0.53	0.77	1.00	0.41	0.46	0.63	0.63	1.00	0.71	0.36	0.40		1.00	0.41	0.53	0.63	0.77	1.00	0.71
14	0.40	0.53	0.77	1.00	0.41	0.46	0.63	0.63	1.00	0.71	0.36	0.40	1.00		0.41	0.53	0.63	0.77	1.00	0.71
15	0.17	0.22	0.32	0.41	1.00	0.19	0.26	0.26	0.41	0.58	0.15	0.17	0.41	0.41		0.22	0.26	0.32	0.41	0.58
16	0.76	1.00	0.69	0.53	0.22	0.86	0.84	0.84	0.53	0.38	0.68	0.76	0.53	0.53	0.22		0.84	0.69	0.53	0.38
17	0.64	0.84	0.81	0.63	0.26	0.73	1.00	1.00	0.63	0.44	0.57	0.64	0.63	0.63	0.26	0.84		0.81	0.63	0.44
18	0.52	0.69	1.00	0.77	0.32	0.59	0.81	0.81	0.77	0.55	0.47	0.52	0.77	0.77	0.32	0.69	0.81		0.77	0.55
19	0.40	0.53	0.77	1.00	0.41	0.46	0.63	0.63	1.00	0.71	0.36	0.40	1.00	1.00	0.41	0.53	0.63	0.77		0.71
20	0.29	0.38	0.55	0.71	0.58	0.32	0.44	0.44	0.71	1.00	0.26	0.29	0.71	0.71	0.58	0.38	0.44	0.55	0.71	

Fig. S1. Overlap calculations (part 1)

Gender overlaps ( $O^g_{ij}$ )																				
Indiv. $ i & j \rightarrow$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		0	1	1	0	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0
2	0		0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1
3	1	0		1	0	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0
4	1	0	1		0	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0
5	0	1	0	0		0	0	0	0	1	0	0	0	1	0	0	0	0	0	1
6	1	0	1	1	0		1	1	1	0	1	1	1	0	1	1	1	1	1	0
7	1	0	1	1	0	1		1	1	0	1	1	1	0	1	1	1	1	1	0
8	1	0	1	1	0	1	1		1	0	1	1	1	0	1	1	1	1	1	0
9	1	0	1	1	0	1	1	1		0	1	1	1	0	1	1	1	1	1	0
10	0	1	0	0	1	0	0	0	0		0	0	0	1	0	0	0	0	0	1
11	1	0	1	1	0	1	1	1	1	0		1	1	0	1	1	1	1	1	0
12	1	0	1	1	0	1	1	1	1	0	1		1	0	1	1	1	1	1	0
13	1	0	1	1	0	1	1	1	1	0	1	1		1	1	1	1	1	1	0
14	0	1	0	0	1	0	0	0	0	1	0	0	0		0	0	0	0	0	1
15	1	0	1	1	0	1	1	1	1	0	1	1	1	0		1	1	1	1	0
16	1	0	1	1	0	1	1	1	1	0	1	1	1	0	1		1	1	1	0
17	1	0	1	1	0	1	1	1	1	0	1	1	1	0	1	1		1	1	0
18	1	0	1	1	0	1	1	1	1	0	1	1	1	0	1	1	1		1	0
19	1	0	1	1	0	1	1	1	1	0	1	1	1	0	1	1	1	1		0
20	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	

Ethnicity overlaps ( $O^e_{ij}$ )																				
Indiv. $ i & j \rightarrow$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		1	0	0	0	0	0	1	1	0	1	1	0	0	1	1	0	0	0	0
2	1		0	0	0	0	0	1	1	0	1	1	0	0	1	1	0	0	0	0
3	0	0		1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
4	0	0	1		0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
5	0	0	0	0		1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
6	0	0	0	0	1		0	0	0	0	0	0	0	0	0	0	0	1	0	0
7	0	0	0	0	0	0		0	0	0	0	0	0	1	0	0	1	0	0	0
8	1	1	0	0	0	0	0		1	0	1	1	0	0	1	1	0	0	0	0
9	1	1	0	0	0	0	0	1		0	1	1	0	0	1	1	0	0	0	0
10	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0
11	1	1	0	0	0	0	0	1	1	0		1	0	0	1	1	0	0	0	0
12	1	1	0	0	0	0	0	1	1	0	1		0	0	1	1	0	0	0	0
13	0	0	1	1	0	0	0	0	0	0	0	0		0	0	0	0	0	0	1
14	0	0	0	0	0	0	1	0	0	0	0	0	0		0	1	0	0	0	0
15	1	1	0	0	0	0	0	1	1	0	1	1	0	0		1	0	0	0	0
16	1	1	0	0	0	0	0	1	1	0	1	1	0	0	1		0	0	0	0
17	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0		0	0	0
18	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0		0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		0
20	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	

Team tenure overlaps ( $O^t_{ij}$ )																				
Indiv. $ i & j \rightarrow$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		0.80	0.40	0.20	0.20	0.40	0.60	0.60	0.08	0.04	0.63	0.71	0.60	0.20	0.20	0.80	0.60	0.40	0.40	0.40
2	0.80		0.50	0.25	0.25	0.50	0.75	0.75	0.10	0.05	0.50	0.57	0.75	0.25	0.25	1.00	0.75	0.50	0.50	0.50
3	0.40	0.50		0.50	0.50	1.00	0.67	0.67	0.20	0.10	0.25	0.29	0.67	0.50	0.50	0.50	0.67	1.00	1.00	1.00
4	0.20	0.25	0.50		1.00	0.50	0.33	0.33	0.40	0.20	0.13	0.14	0.33	1.00	1.00	0.25	0.33	0.50	0.50	0.50
5	0.20	0.25	0.50	1.00		0.50	0.33	0.33	0.40	0.20	0.13	0.14	0.33	1.00	1.00	0.25	0.33	0.50	0.50	0.50
6	0.40	0.50	1.00	0.50	0.50		0.67	0.67	0.20	0.10	0.25	0.29	0.67	0.50	0.50	0.50	0.67	1.00	1.00	1.00
7	0.60	0.75	0.67	0.33	0.33	0.67		1.00	0.13	0.07	0.38	0.43	1.00	0.33	0.33	0.75	1.00	0.67	0.67	0.67
8	0.60	0.75	0.67	0.33	0.33	0.67	1.00		0.13	0.07	0.38	0.43	1.00	0.33	0.33	0.75	1.00	0.67	0.67	0.67
9	0.08	0.10	0.20	0.40	0.40	0.20	0.13	0.13		0.50	0.05	0.06	0.13	0.40	0.40	0.10	0.13	0.20	0.20	0.20
10	0.04	0.05	0.10	0.20	0.20	0.10	0.07	0.07	0.50		0.03	0.03	0.07	0.20	0.20	0.05	0.07	0.10	0.10	0.10
11	0.63	0.50	0.25	0.13	0.13	0.25	0.38	0.38	0.05	0.03		0.88	0.38	0.13	0.13	0.50	0.38	0.25	0.25	0.25
12	0.71	0.57	0.29	0.14	0.14	0.29	0.43	0.43	0.06	0.03	0.88		0.43	0.14	0.14	0.57	0.43	0.29	0.29	0.29
13	0.60	0.75	0.67	0.33	0.33	0.67	1.00	1.00	0.13	0.07	0.38	0.43		0.33	0.33	0.75	1.00	0.67	0.67	0.67
14	0.20	0.25	0.50	1.00	1.00	0.50	0.33	0.33	0.40	0.20	0.13	0.14	0.33		1.00	0.25	0.33	0.50	0.50	0.50
15	0.20	0.25	0.50	1.00	1.00	0.50	0.33	0.33	0.40	0.20	0.13	0.14	0.33	1.00		0.25	0.33	0.50	0.50	0.50
16	0.80	1.00	0.50	0.25	0.25	0.50	0.75	0.75	0.10	0.05	0.50	0.57	0.75	0.25	0.25		0.75	0.50	0.50	0.50
17	0.60	0.75	0.67	0.33	0.33	0.67	1.00	1.00	0.13	0.07	0.38	0.43	1.00	0.33	0.33	0.75		0.67	0.67	0.67
18	0.40	0.50	1.00	0.50	0.50	1.00	0.67	0.67	0.20	0.10	0.25	0.29	0.67	0.50	0.50	0.50	0.67		1.00	1.00
19	0.40	0.50	1.00	0.50	0.50	1.00	0.67	0.67	0.20	0.10	0.25	0.29	0.67	0.50	0.50	0.50	0.67	1.00		1.00
20	0.40	0.50	1.00	0.50	0.50	1.00	0.67	0.67	0.20	0.10	0.25	0.29	0.67	0.50	0.50	0.50	0.67	1.00	1.00	

Fig. S1. Overlap calculations (part 2)