

Group Projects

Bamberg Summer Institute in Computational Social Science

Carsten Schwemmer, University of Bamberg

2019-08-05

Many thanks to Chris Bail for providing material for this exercise

Strengths of group projects

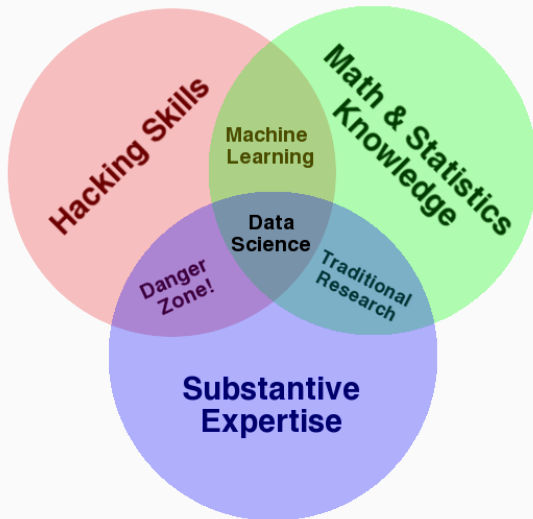
Atypical Combinations and Scientific Impact

Brian Uzzi,^{1,2} Satyam Mukherjee,^{1,2} Michael Stringer,^{2,3} Ben Jones^{1,4*}

Novelty is an essential feature of creative ideas, yet the building blocks of new ideas are often embodied in existing knowledge. From this perspective, balancing atypical knowledge with conventional knowledge may be critical to the link between innovativeness and impact. Our analysis of 17.9 million papers spanning all scientific fields suggests that science follows a nearly universal pattern: The highest-impact science is primarily grounded in exceptionally conventional combinations of prior work yet simultaneously features an intrusion of unusual combinations. Papers of this type were twice as likely to be highly cited works. Novel combinations of prior work are rare, yet teams are 37.7% more likely than solo authors to insert novel combinations into familiar knowledge domains.

Efficient way to learn from each other

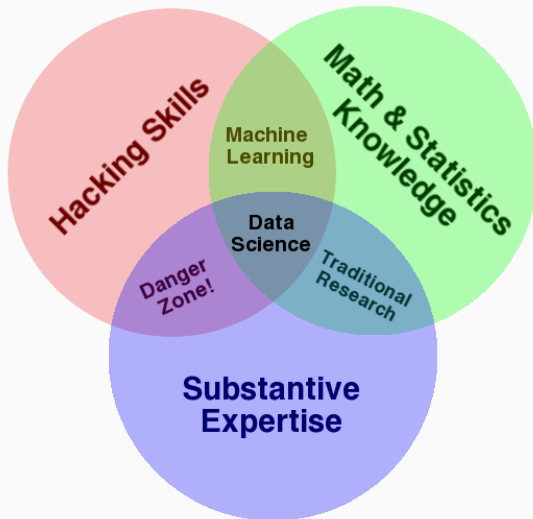


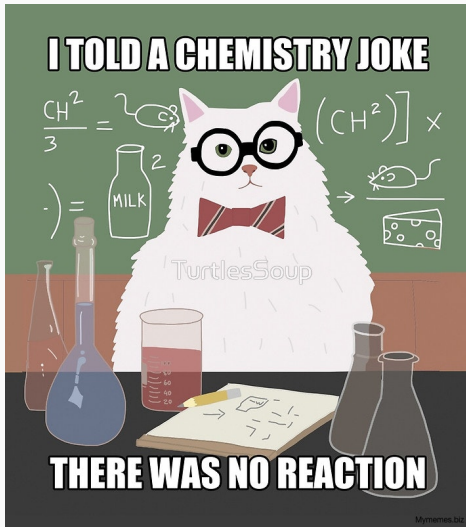


A lot of fun



Weaknesses of group projects

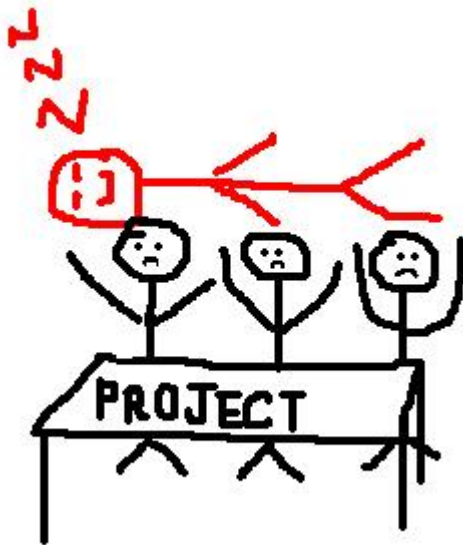






Different goals?





- can take on many forms (from original empirical research to creation of open-source tools)
- many groups will only find enough time to create a proposal; some may have pilot results by the end of the week.
- limited seed funding is available for pilot research, data purchasing, and/or cloud computing costs (we encourage you to explore options such as AWS educate first).

Research speed dating

1. we crowdsource a list of research interests in a google sheet (5 min)
2. each person enters a “1” under their research interests (5 min)
3. we identify maximally similar clusters of participants
4. we identify maximally different clusters of participants

1. maximally similar clusters come up with a group project (30 min)
2. maximally different clusters come up with a group project (30 min)
3. we start a new google doc. Each person takes 5 minutes to write down their favorite group project idea.
4. we have a 20-30 minute discussion of ideas in the google doc with all SICSS participants.
5. everyone puts their name next to the group project they want to join

Timeline for group projects

Monday afternoon: refine group project idea

Tuesday morning: write 1 page project description

Tuesday afternoon: organizers will begin responding to seed funding requests (if there are any) on a rolling basis

Tuesday afternoon-Thursday morning: work on group projects

Thursday afternoon/evening: group presentations (with feedback)

Required deliverable (for Thursday)

A document that contains a presentation of your group project that is between 10-20 minutes that explains:

- why your group project is important
- what are your hypotheses or research questions?
- what data will/did you collect?
- first results (if available)
- what are the next steps?

Note: not all projects will work; if yours fails, please write a post-mortem that explains why

Let's begin

Let's crowdsource a list of our research interests

- navigate to the “Bamberg” tab of the Google Sheet:
<https://bit.ly/2WIrIYh>
- create a new row with your name in the first column and a **1** under each of your research interests

Reading the Google Sheet

```
# Read in interests from google sheet
library(googleheets)
library(tidyverse)
library(caTools)

SICSS_sheet <- gs_url("https://docs.google.com/spreadsheets/d/1RIeYYvtzEX2kk

# Extract interests for site
interests <- gs_read(SICSS_sheet,
                     ws = "Bamberg") # will be replaced with Bamberg
interests <- interests %>%
  select(-c(2,3)) %>%
  slice(2:nrow(interests))
```

Identify maximally similar groups

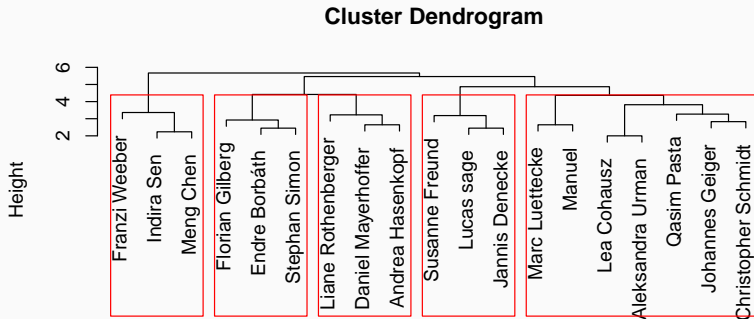
```
# Create distance matrix
interests <- interests %>%
  # turn interest vectors into numeric
  mutate_at(vars(-Name), as.numeric) %>%
  # replace nas with 0
  replace(is.na(.), 0)

interest_mat <- interests %>%
  column_to_rownames("Name") %>%
  as.matrix()

sim_matrix <- dist(interest_mat, method = "euclidean")
fit_interests <- hclust(sim_matrix , method = "ward.D")
```

Identify maximally similar groups

```
# Display dendrogram  
plot(fit_interests)  
# Draw dendrogram with cut points for groups  
rect.hclust(fit_interests, k = 5, border = "red")
```



```
sim_matrix  
hclust (*, "ward.D")
```

Compute distances for random groups

```
# Identify all unique combinations of participants of size 4
all_combs <- combs(interests$Name, 4)
# Calculate mean distance within groups
dissimilar_groups <- as.data.frame(all_combs)
dissimilar_groups$score <- NA
for (i in 1:nrow(all_combs)) {
  temp_group <- interests[interests$Name %in% all_combs[i,],]
  temp_dist <- mean(dist(temp_group[,-1]))
  dissimilar_groups$score[i] <- temp_dist
}
```


Identify maximally diverse groups

```
dis_groups <- max_diverse(dissimilar_groups)
dis_groups
```

```
##                Group 1          Group 2          Group 3
## Member 1      Marc Luetetecke  Endre Borbáth      Manuel
## Member 2      Indira Sen       Susanne Freund   Qasim Pasta
## Member 3      Daniel Mayerhoffer  Franzi Weeber   Meng Chen
## Member 4      Jannis Denecke     Lucas sage      Andrea Hasenkopf
##                Group 4                Group 5
## Member 1      Johannes Geiger      Lea Cohausz
## Member 2      Stephan Simon        Aleksandra Urman
## Member 3      Liane Rothenberger  Christopher Schmidt
## Member 4      Florian Gilberg
```

Let's go!

- Meet in maximally similar and dissimilar groups for 30 minutes.
- At the end of each 30 minute period, one group representative should write the name of the project and a brief description. Here's a Google Sheet for project ideas: https://docs.google.com/spreadsheets/d/17n2flcUlsG8oXDp0m0b73lc9Z89_JnYNFXB_W5pE1qk/edit?usp=sharing
- After the end of the exercise, there will be 5 more minutes where everyone can add more project ideas to the sheet.
- Put your name next to the research project that you are most excited about joining.
- Have lunch with your group to begin discussing your project.
- It is ok to change groups until the end of today.

- Monday 12.30 - 13.30: lunch
- Monday 13.30 - 15.00: guest talk by Martijn Schoonvelde
- Tuesday 12.30 - 13.30: lunch
- Tuesday 13.30 - 15.00: guest talk by Milena Tsvetkova
- Wednesday 12.30 - 13.30: lunch
- Wednesday 13.30 - 15.00: flash talks (message organizers on Slack)

Questions?