

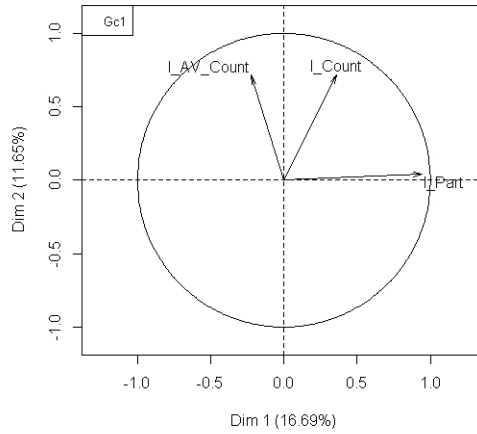
# **Data Quality of Digital Process Data**

## **A Generalized Framework and Simulation/Post-Hoc-Identification Strategy**

**Andreas Schmitz · Jan R. Riebling**

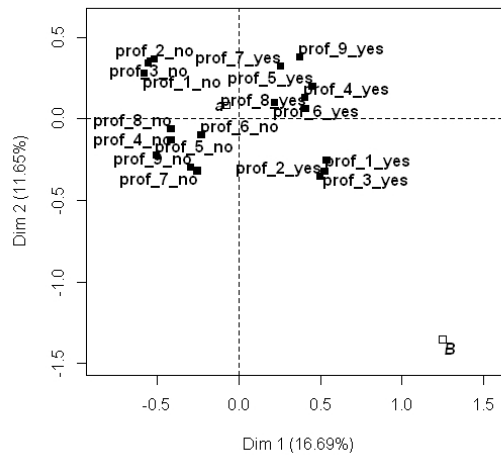
**Online Appendix**

**Fig. A1** Dataset 1 – Continuous Variables (Biplot)

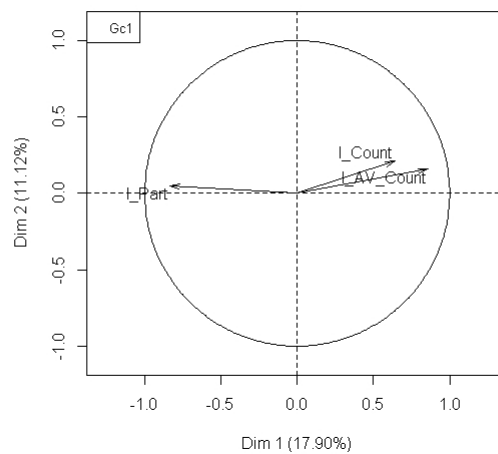


*I\_Count* = number of overall interactions; *I\_AV\_Count* = average number of interactions per contact;  
*I\_Part* = number of distinct partners

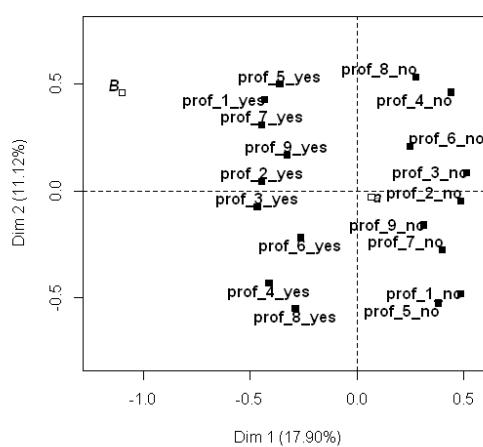
**Fig. A2** Dataset 1 – Categorical Variables



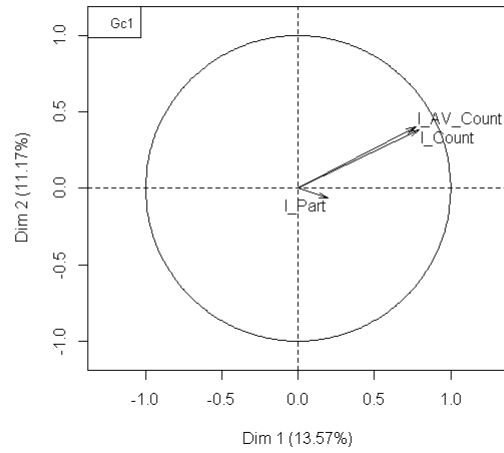
*Profile-indicators (1-9)*; *a*= average position of non-Bots; *B*= average bot position

**Fig. B1** Dataset 2 – Continuous Variables (Biplot)

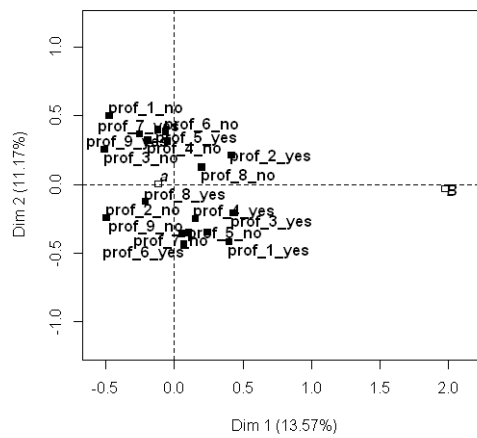
*I\_Count* = number of overall interactions; *I\_AV\_Count* = average number of interactions per contact;  
*I\_Part* = number of distinct partners

**Fig. B2** Dataset 2 – Categorical Variables

*Profile-indicators (1-9)*; *a* = average position of non-Bots; *B* = average bot position

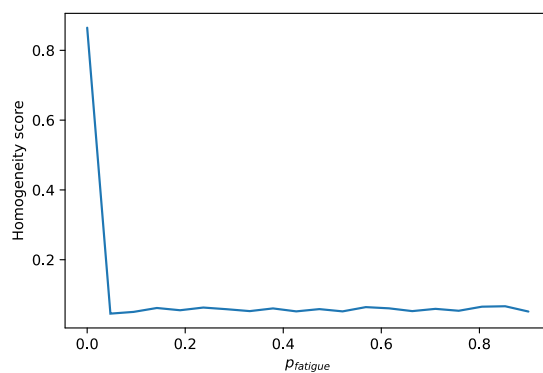
**Fig. C1** Dataset 3 – Continuous Variables (Biplot)

$I\_Count$  = number of overall interactions;  $I\_AV\_Count$  = average number of interactions per contact;  
 $I\_Part$  = number of distinct partners

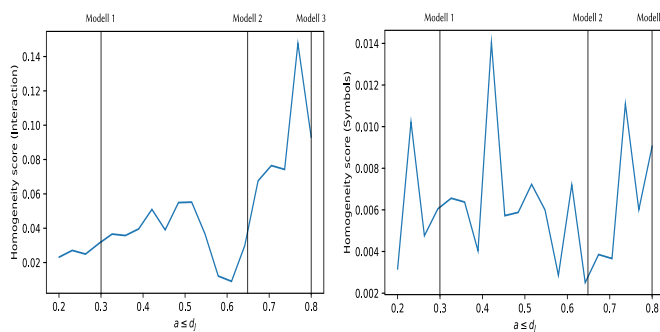
**Fig. C2** Dataset 3 – Categorical Variables

Profile-indicators (1-9); a= average position of non-Bots; B= average bot position

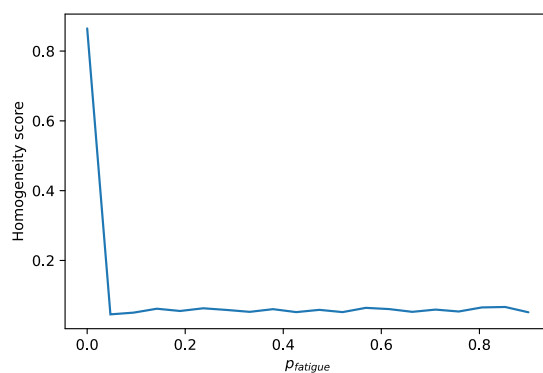
**Fig. D1** Homogeneity score for k-means clusters ( $k=2$ ) for different values of the base probability of a bot terminating an interaction.



**Fig. D2** Homogeneity scores for k-means clusters ( $k=2$ ) for different values of  $a$ .



**Fig. D3** Homogeneity score for k-means clusters ( $k=2$ ) for different values of the base probability of a bot terminating an interaction.



**Fig. D4** Homogeneity scores for k-means clusters (k=2) for different values of a.

