Application of Significance Tests to Massive Open Online Courses (MOOCs)

Simon Coton

Steve Cayzer

Background

- Massive Open Online Courses (MOOC's) first introduced in 2008 [1]
- University of Bath MOOCs – "Inside Cancer", "Sustainability" - 11 runs
  - Chance to compare different styles of MOOC
  - Chance to look at network dynamics within and between runs

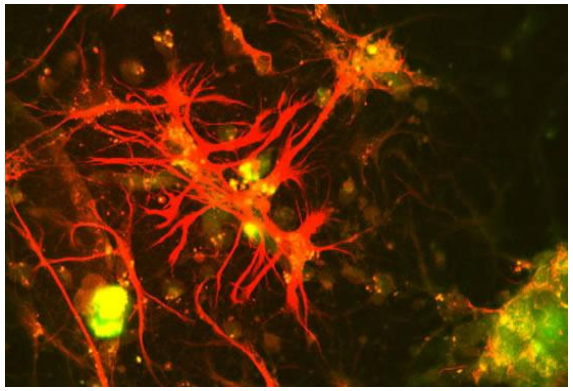


Figure 1: Inside Cancer Course

Figure 2: Sustainability Course

Background

- Significance tests distinguish whether a relationship is significant or has appeared by chance
- Initial literature revealed significance tests used inconsistently
  - Do not quantify significance [2][3][4]
  - T-test assumed [5][6]
- This creates a problem – credibility of results

Network Cohesiveness

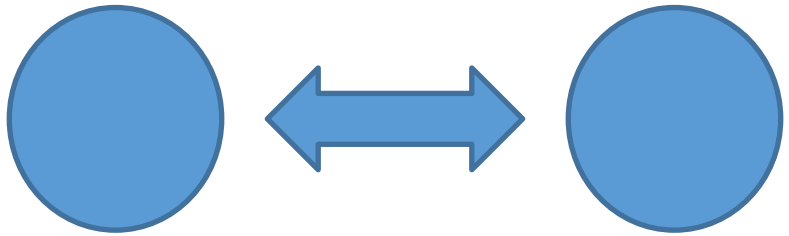
“reciprocity and transitivity … [create] cohesive clusters, which support … different information sources …[providing] students with multiple channels to access information and knowledge … To increase the level of reciprocity and transitivity in scalable discussions seems like a useful strategy”

(although)

“.. an increased level of cohesiveness .. [leads to] … network closure … participants are likely to become more selective when interacting with others”

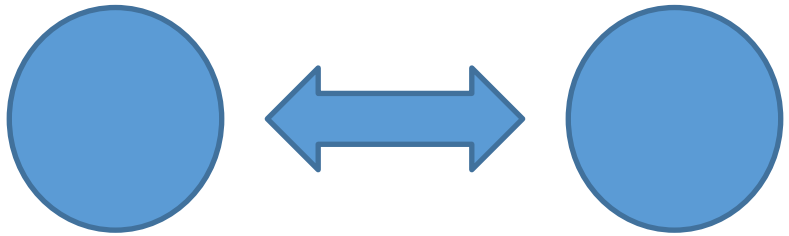
Zhang et al (2016).

Reciprocity

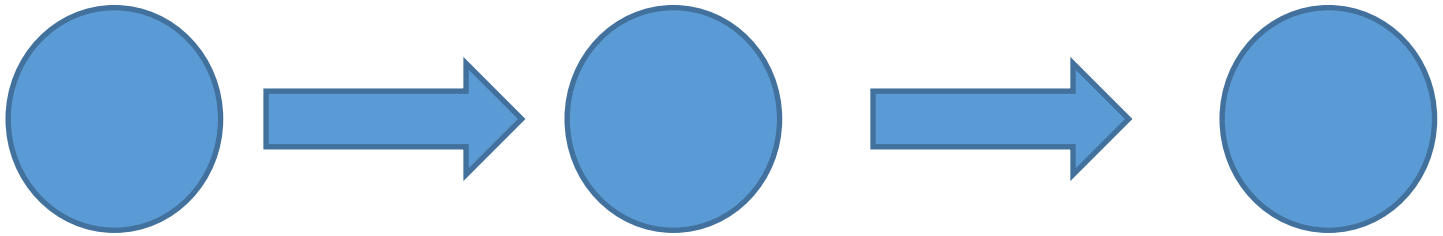


Reciprocity

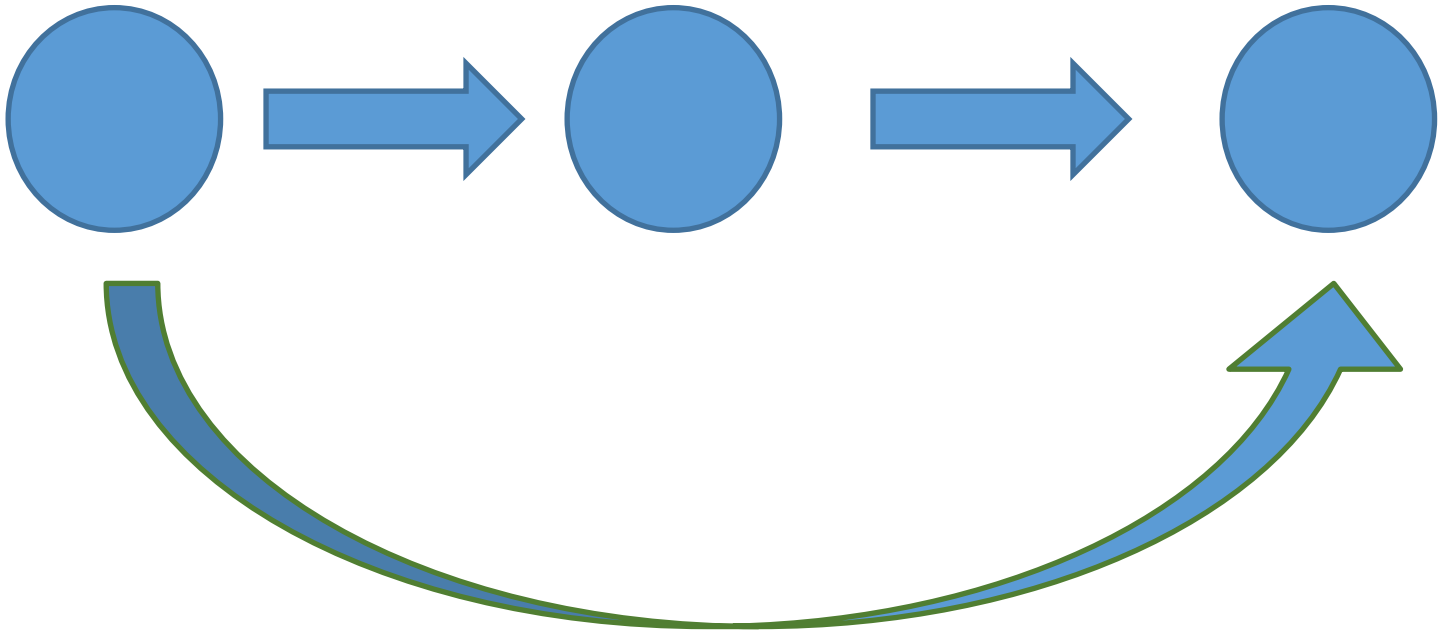
I talk to you, you talk to me



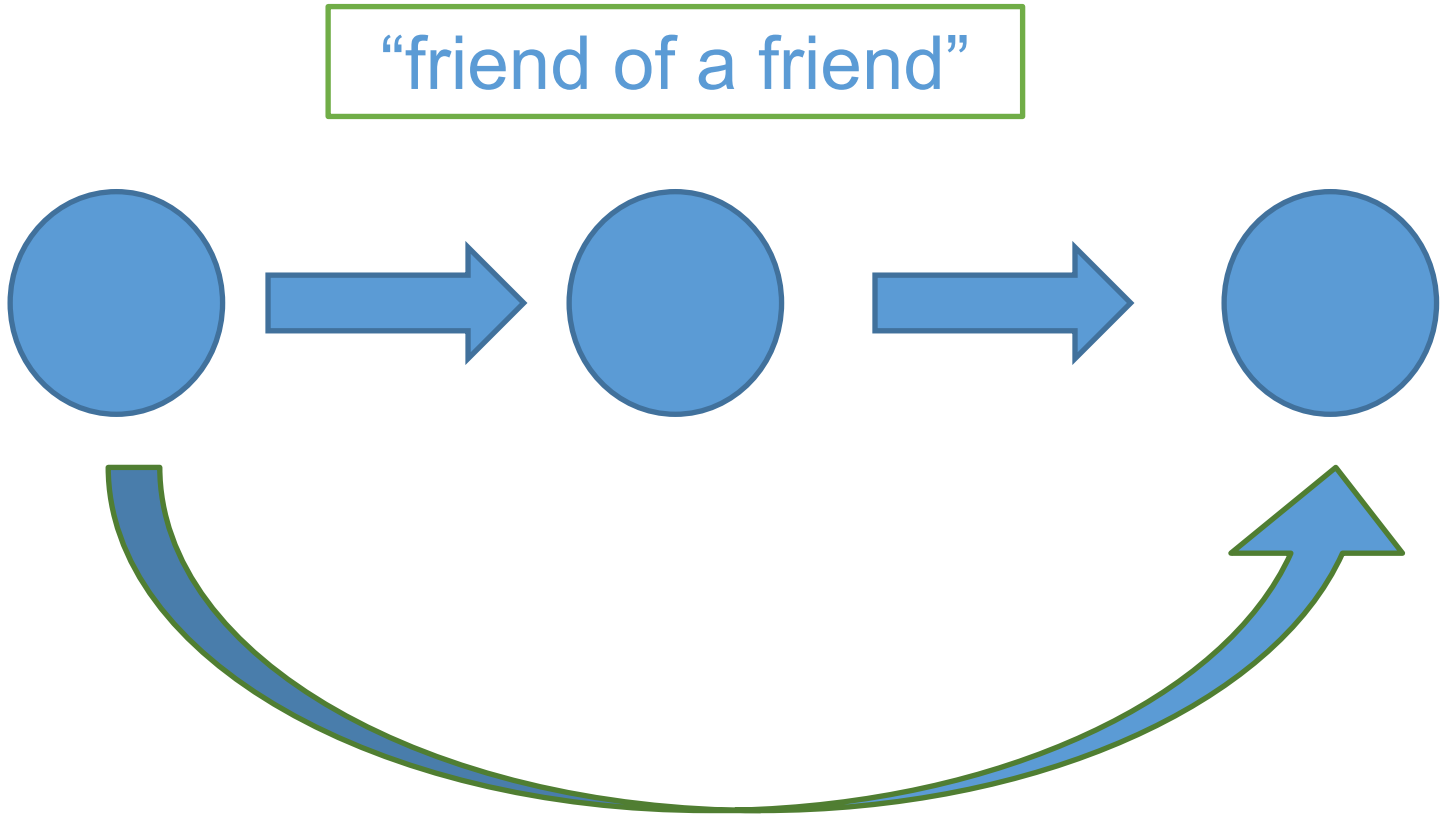
Transitivity



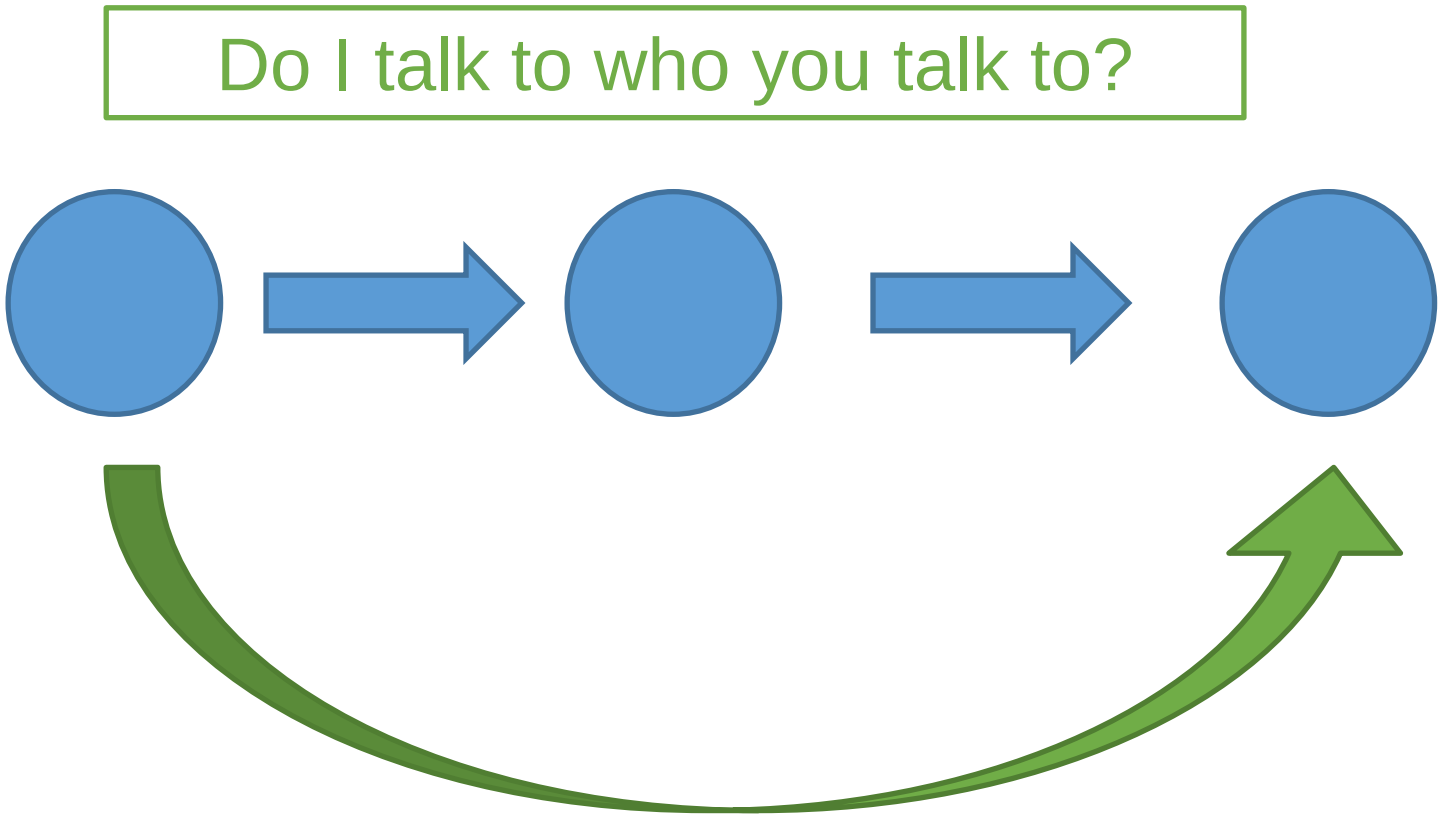
Transitivity



Transitivity



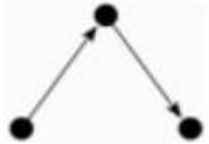
Transitivity



Transitivity

012

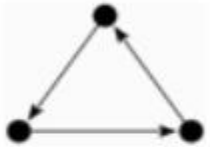
102



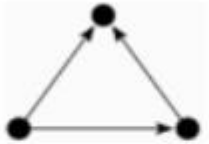
021C

021D

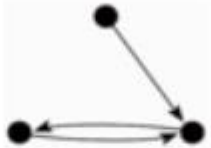
021U



030C



030T



111D

111U

120C

120D

120U

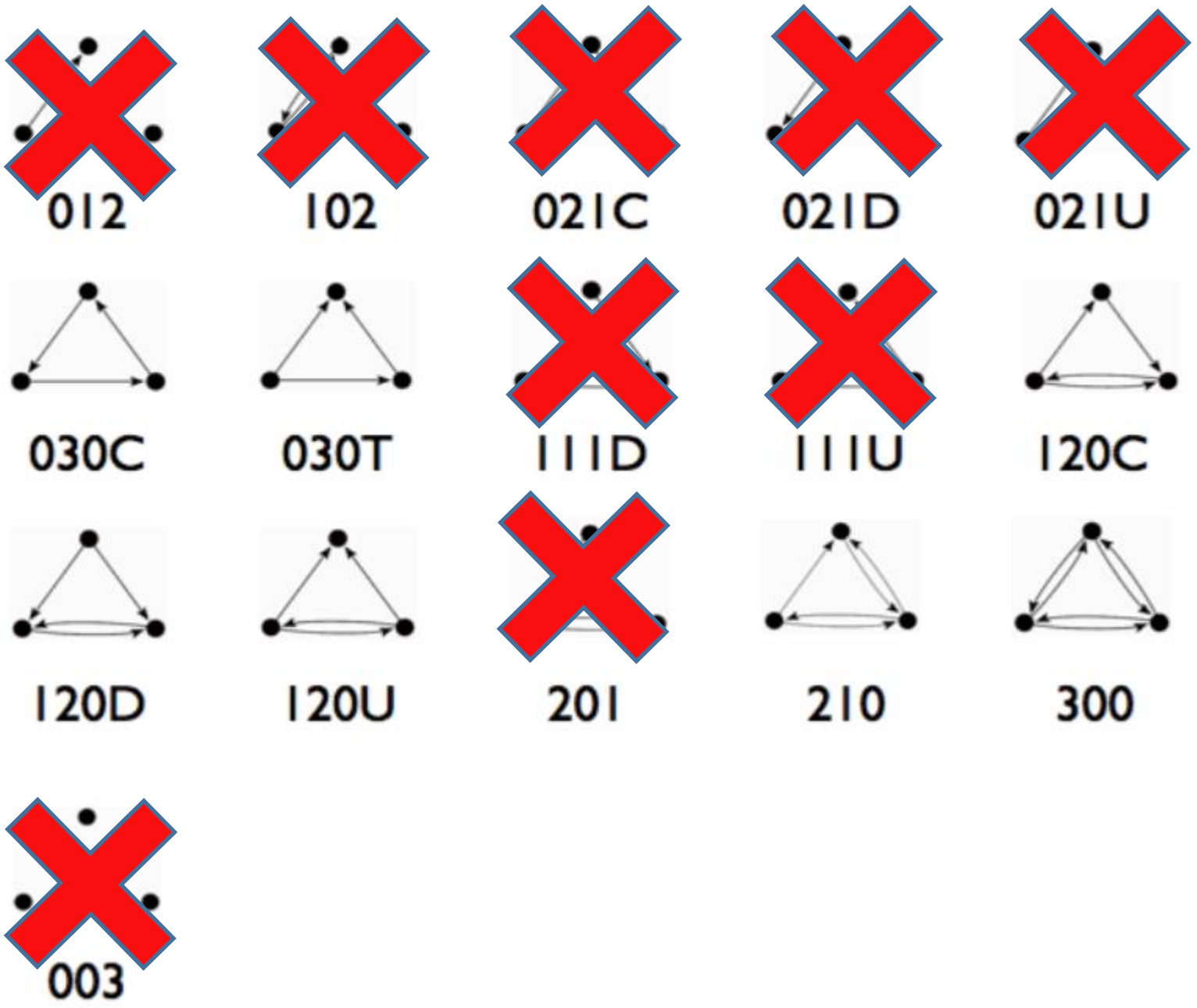
201

210

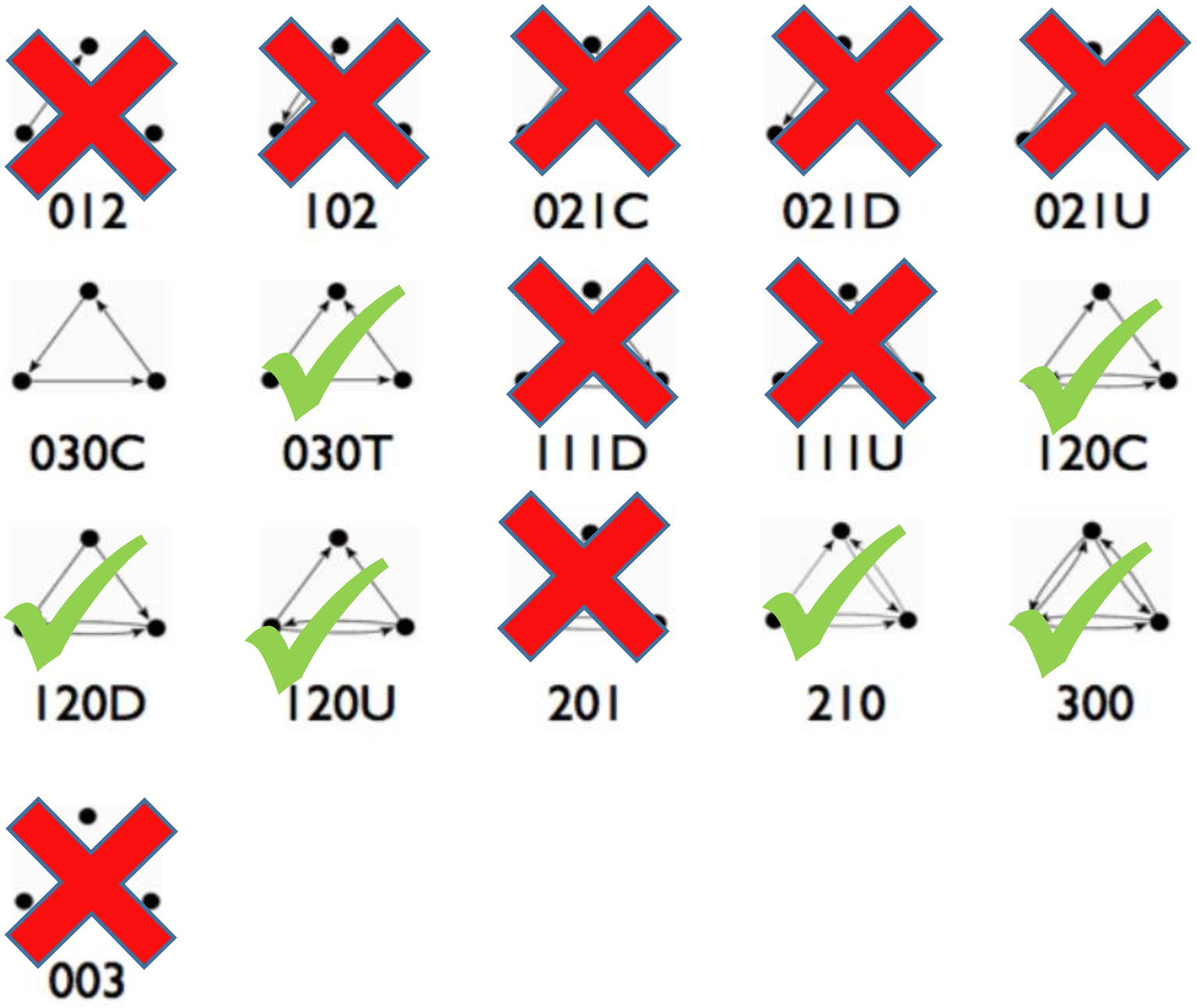
300

003

Transitivity

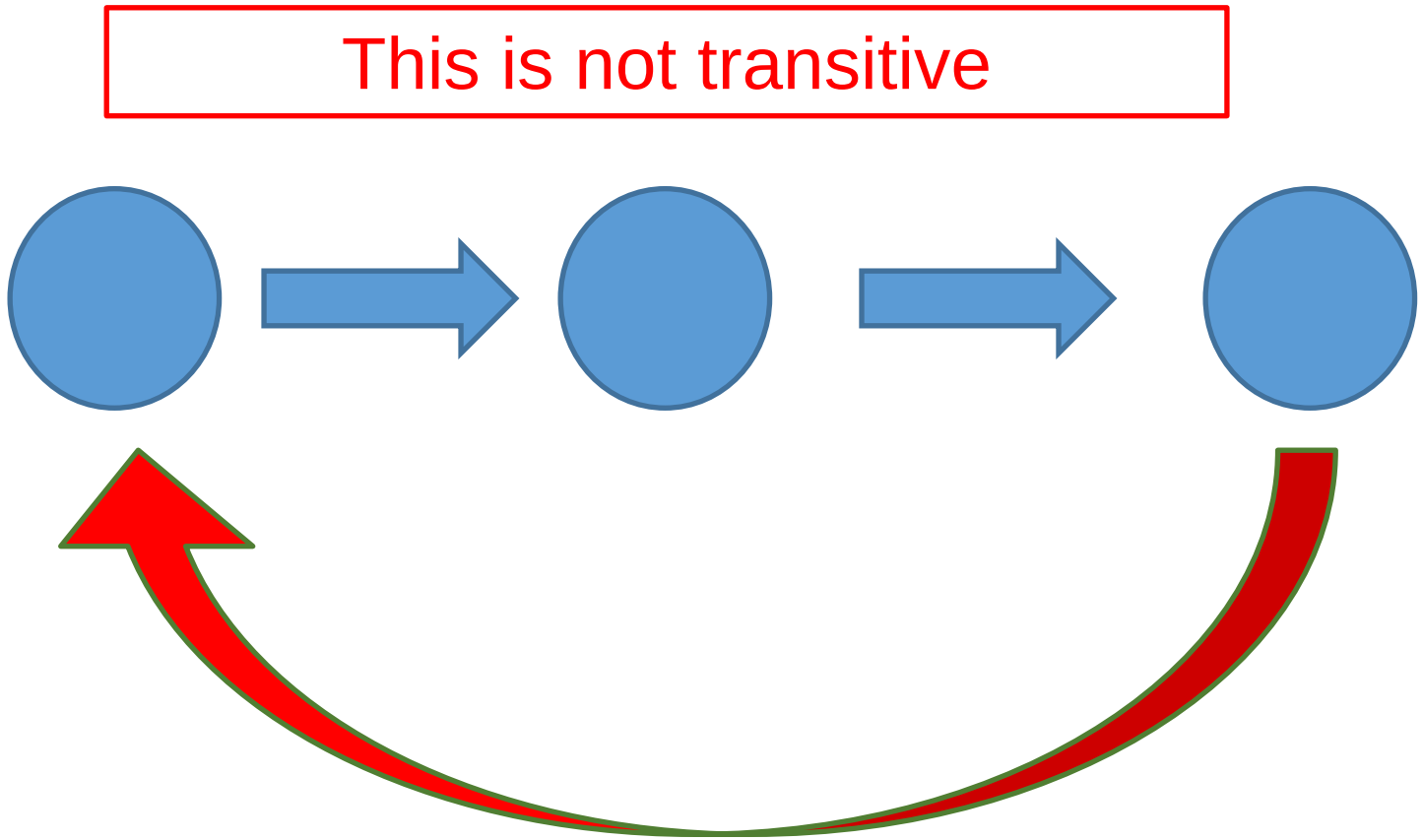


Transitivity



Transitivity

Transitivity



Transitivity

Stats Tests

Pearson's Correlation

Assumes normality

strength and direction of a **linear relationship**

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

x_i = value

Spearman's Rank

non-parametric equivalent

strength and direction of a **monotonic relationship**

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

x_i = rank

https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php

Results-Reciprocity

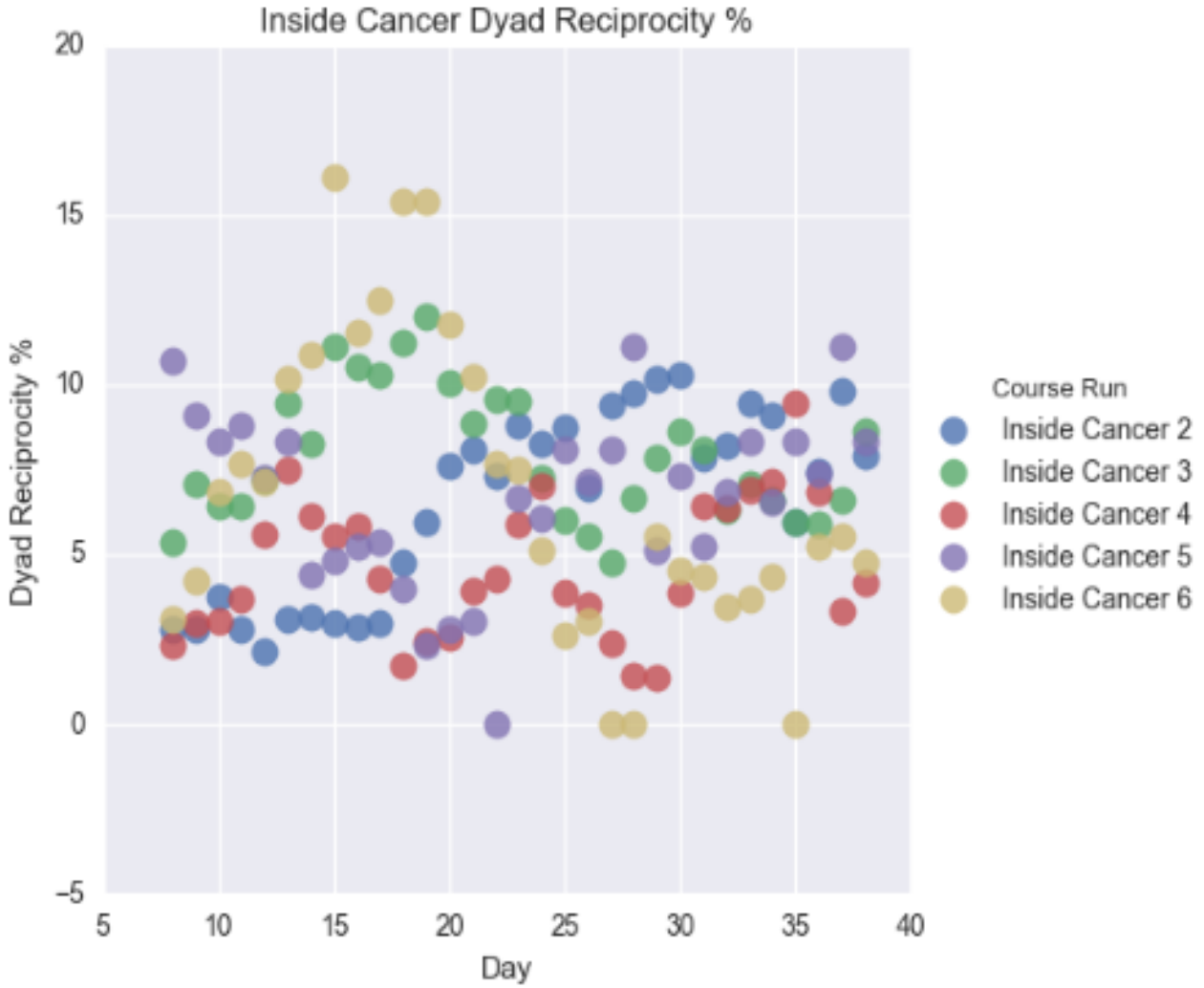


Figure 3: Rolling average of reciprocity over time for Inside Cancer runs 2-6

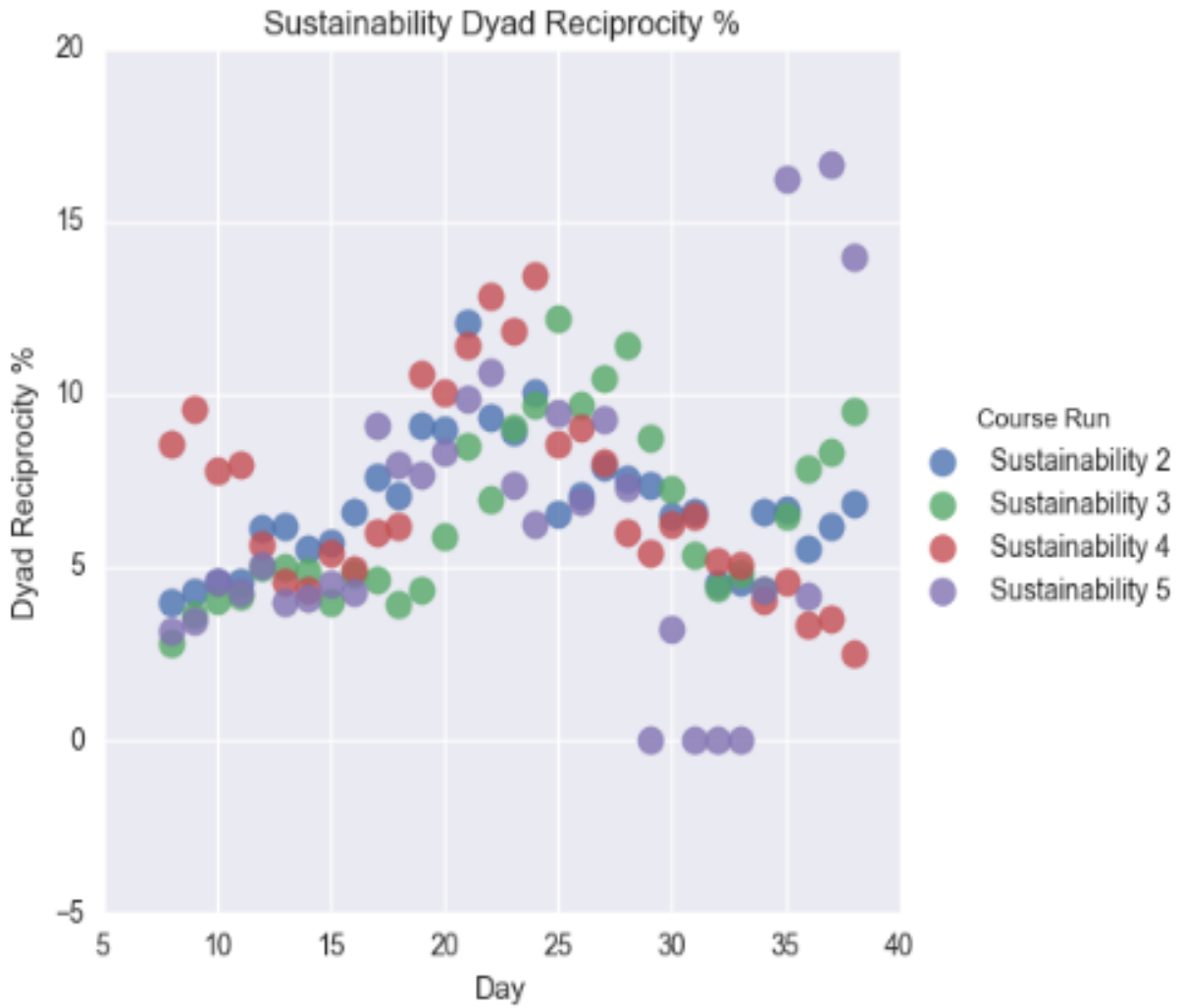


Figure 4: Rolling average of reciprocity over time for Sustainability runs 2-5

Results-Transitivity

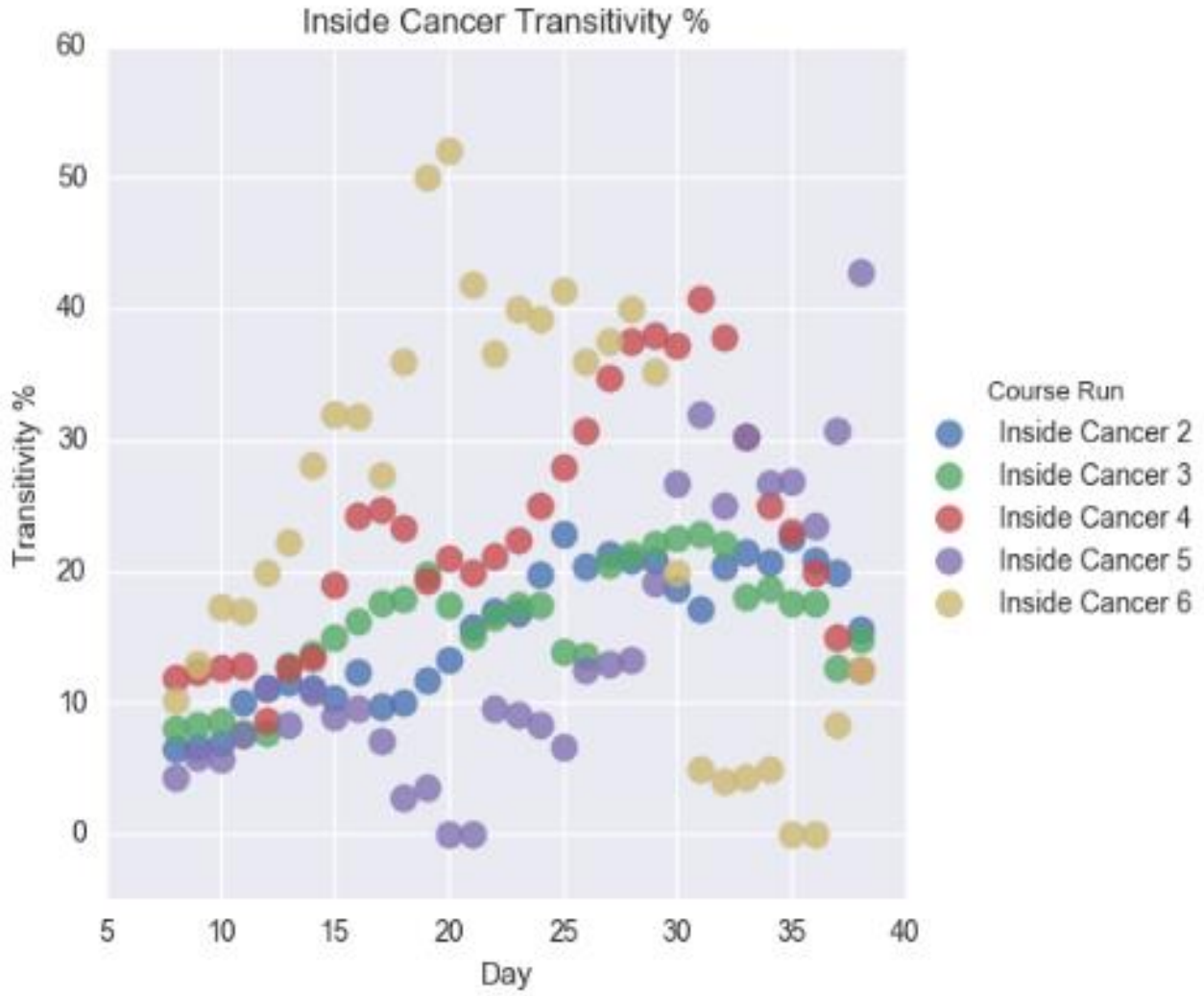


Figure 5: Rolling average of transitivity over time for Inside Cancer runs 2-6

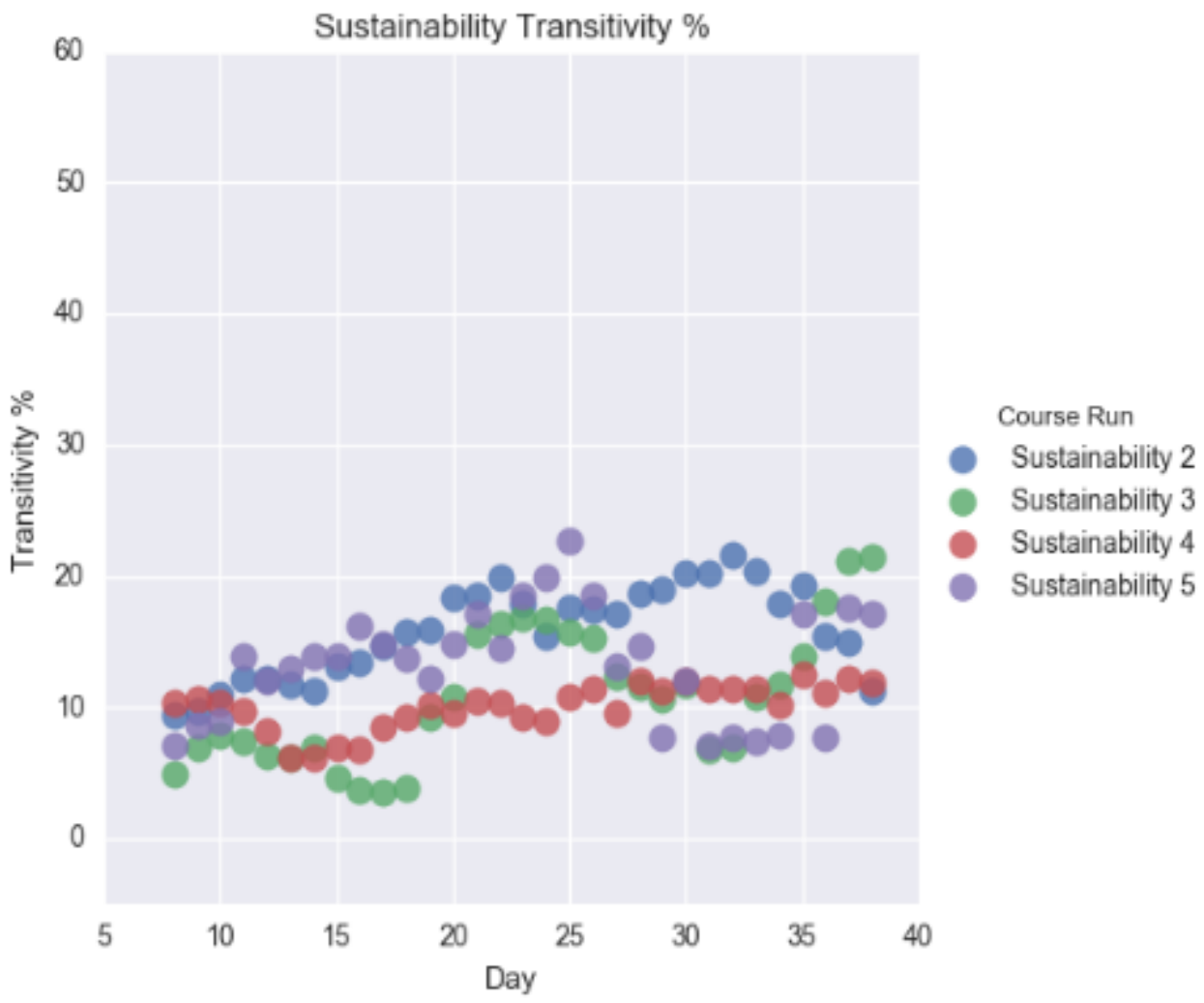
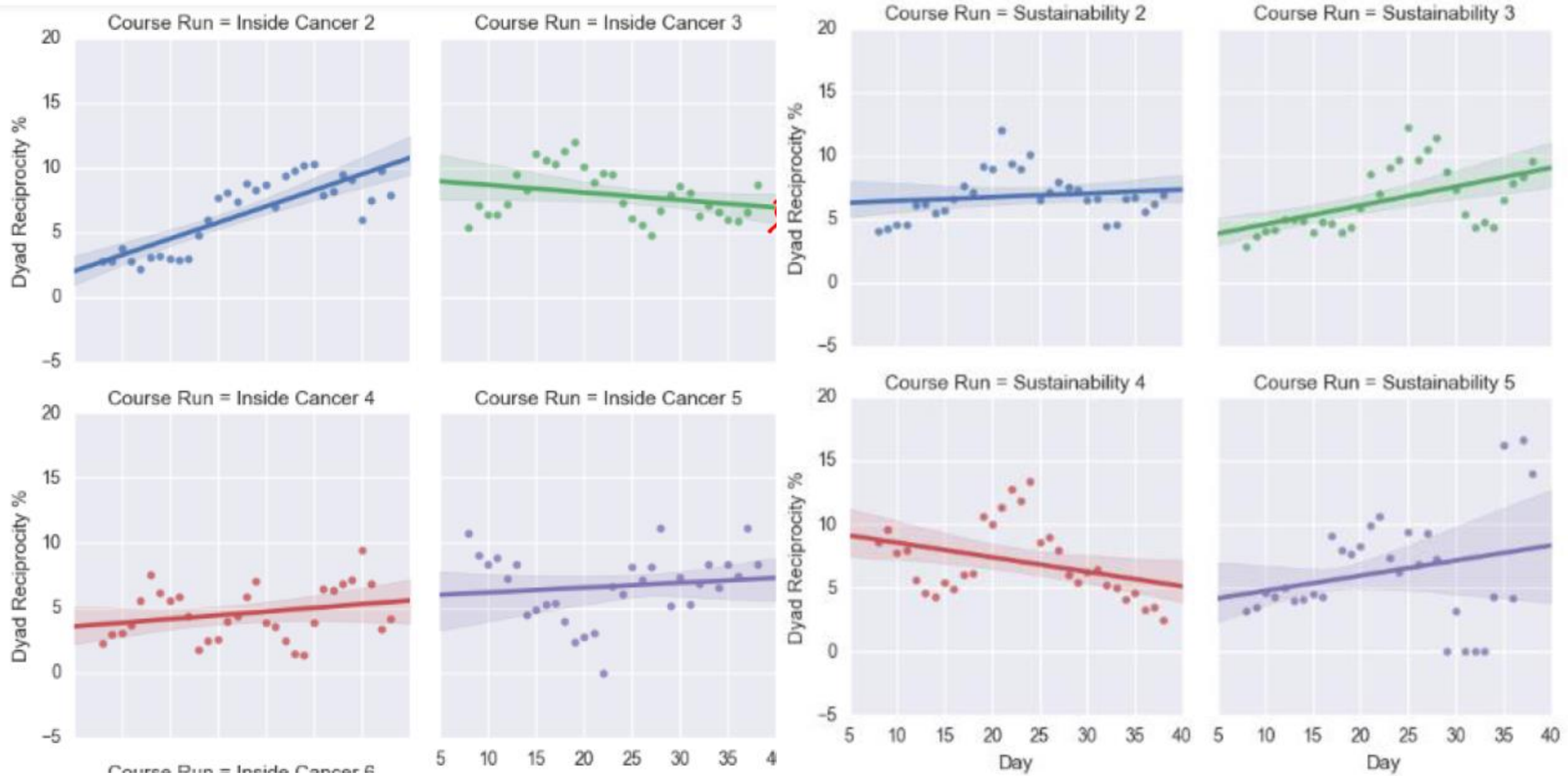
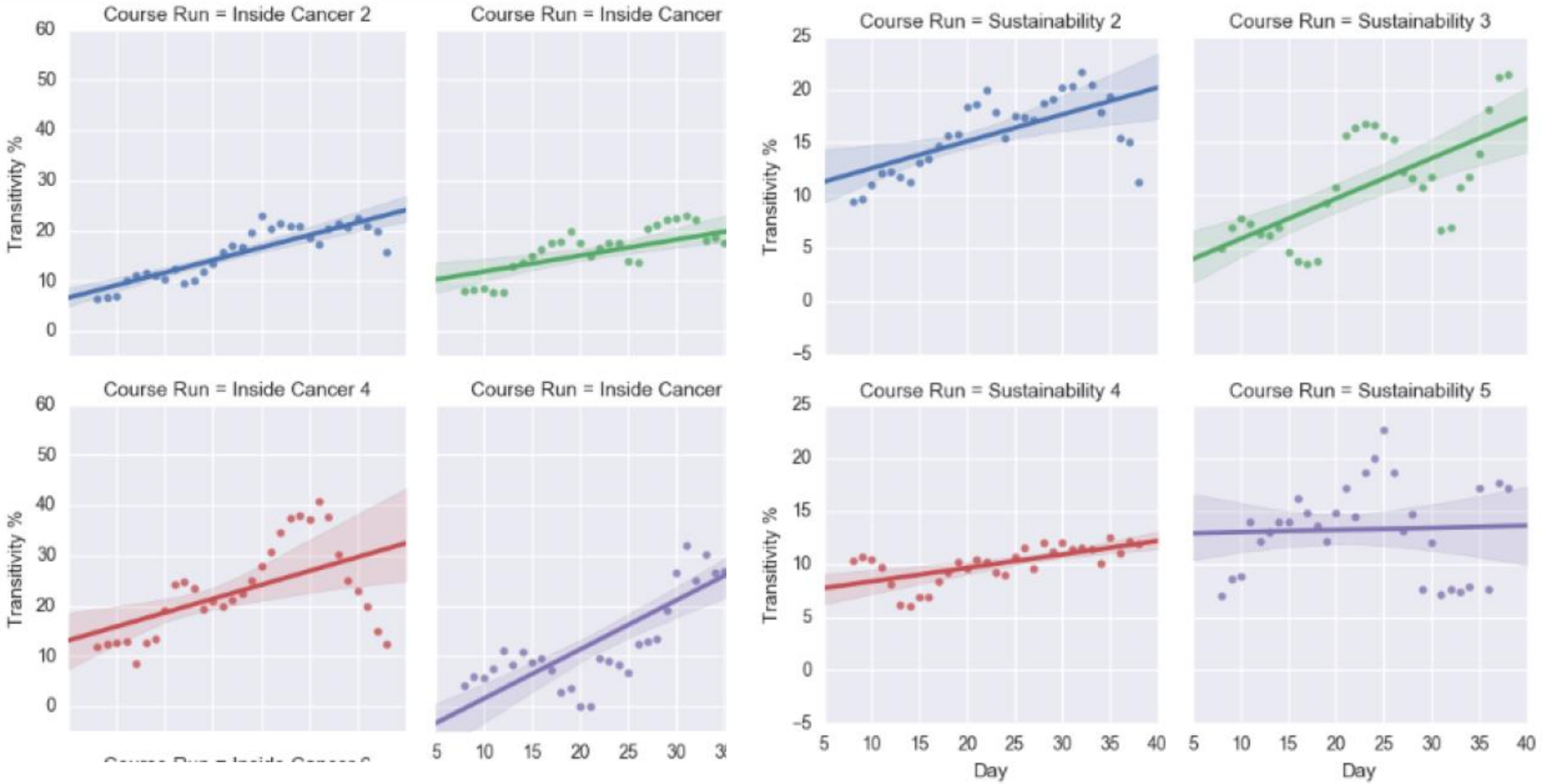


Figure 6: Rolling average of transitivity over time for Sustainability runs 2-5

Reciprocity



Transitivity



Reciprocity

Key

- Not significant at 0.05, 0.01 and 0.001 levels
- Significant at 0.05 level but fails to be conclusive at 0.01 and 0.001 levels
- Significant at 0.05 and 0.01 levels but fails to be conclusive at 0.001 level
- Significant at 0.05, 0.01 and 0.001 levels

Course Run	Spearman's Correlation	Spearman's p-value (Single Test)	Permuted Spearman's p-value	Pearson's Correlation	Pearson's p-value (Single Test)	Permuted Pearson's p-value
Combined Course Reciprocity						
Inside Cancer Runs 2-6	0.071	0.378	0.199	0.027	0.743	0.380
Sustainability Runs 2-5	0.143	0.112	0.057	0.136	0.133	0.064
Individual Runs Reciprocity						
Inside Cancer Run 2	0.775	3.00E-07	0.000	0.819	1.83E-08	0.000
Inside Cancer Run 3	-0.261	0.160	0.089	-0.271	0.140	0.070
Inside Cancer Run 4	0.259	0.259	0.086	0.255	0.255	0.008
Inside Cancer Run 5	0.098	0.601	0.294	0.131	0.482	0.244
Inside Cancer Run 6	-0.468	0.008	0.006	-0.492	0.005	0.004
Sustainability Run 2	0.201	0.278	0.140	0.146	0.434	0.224
Sustainability Run 3	0.550	0.001	0.002	0.519	0.003	0.001
Sustainability Run 4	-0.393	0.029	0.013	-0.359	0.047	0.021
Sustainability Run 5	0.171	0.359	0.183	0.252	0.172	0.087

Transitivity

Key

- Not significant at 0.05, 0.01 and 0.001 levels
- Significant at 0.05 level but fails to be conclusive at 0.01 and 0.001 levels
- Significant at 0.05 and 0.01 levels but fails to be conclusive at 0.001 level
- Significant at 0.05, 0.01 and 0.001 levels

Course Run	Spearman's Correlation	Spearman's p-value (Single Test)	Permuted Spearman's p-value	Pearson's Correlation	Pearson's p-value (Single Test)	Permuted Pearson's p-value
Combined Course Reciprocity						
Inside Cancer Runs 2-6	0.378	1.00E-05	0.000	0.303	1.00E-04	0.000
Sustainability Runs 2-5	0.388	8.00E-06	0.000	0.392	7.00E-06	0.000
Individual Runs Reciprocity						
Inside Cancer Run 2	0.821	1.49E-08	0.000	0.867	3.00E-10	0.000
Inside Cancer Run 3	0.640	4.00E-04	0.001	0.642	1.00E-04	0.000
Inside Cancer Run 4	0.557	1.11E-03	0.001	0.544	0.002	0.003
Inside Cancer Run 5	0.780	2.34E-07	0.000	0.812	3.03E-08	0.000
Inside Cancer Run 6	-0.300	0.101	0.052	-0.318	0.081	0.049
Sustainability Run 2	0.633	1.30E-04	0.000	0.654	6.62E-05	0.000
Sustainability Run 3	0.640	1.07E-04	0.000	0.670	4.20E-05	0.000
Sustainability Run 4	0.680	2.54E-05	0.000	0.655	6.46E-05	0.000
Sustainability Run 5	0.069	0.711	0.341	0.043	0.829	0.413

Transitivity

Key

- Not significant at 0.05, 0.01 and 0.001 levels
- Significant at 0.05 level but fails to be conclusive at 0.01 and 0.001 levels
- Significant at 0.05 and 0.01 levels but fails to be conclusive at 0.001 level
- Significant at 0.05, 0.01 and 0.001 levels

Course Run	Spearman's Correlation	Spearman's p-value (Single Test)	Permuted Spearman's p-value	Pearson's Correlation	Pearson's p-value (Single Test)	Permuted Pearson's p-value
Combined Course Reciprocity						
Inside Cancer Runs 2-6	0.378	1.00E-05	0.000	0.303	1.00E-04	0.000
Sustainability Runs 2-5	0.388	8.00E-06	0.000	0.392	7.00E-06	0.000
Individual Runs Reciprocity						
Inside Cancer Run 2	0.821	1.49E-08	0.000	0.867	3.00E-10	0.000
Inside Cancer Run 3	0.640	4.00E-04	0.001	0.642	1.00E-04	0.000
Inside Cancer Run 4	0.557	1.11E-03	0.001	0.544	0.002	0.003
Inside Cancer Run 5	0.780	2.34E-07	0.000	0.812	3.03E-08	0.000
Inside Cancer Run 6	-0.300	0.101	0.052	-0.318	0.081	0.049
Sustainability Run 2	0.633	1.30E-04	0.000	0.654	6.62E-05	0.000
Sustainability Run 3	0.640	1.07E-04	0.000	0.670	4.20E-05	0.000
Sustainability Run 4	0.680	2.54E-05	0.000	0.655	6.46E-05	0.000
Sustainability Run 5	0.069	0.711	0.341	0.043	0.829	0.413

Outcomes and Further Work

Does Network cohesion increase over duration of MOOC run?

Table 1: Results, implications and future work

Result	Outcome	Further Work
Reciprocity	Non-conclusive.	Focus on interactions by topic, rather than by time. Weightings conversations.
Transitivity	Hypothesis accepted.	Focus on other triad measures – 15 remaining types [7].
Significance Tests	Spearman's correlation most appropriate.	Additional hypotheses using different test statistics.

Questions

Thank you

References (1/2)

[1] Downes, S., 2008. *Connectivism and Connective Knowledge.* Available at: http://connect.downes.ca/archive/08/09_15_ thedaily.htm.

[2] Kop, R. & Fournier, H., 2011. A Pedagogy of Abundance or a Pedagogy to Support Human Beings? Participant Support on Massive Open Online Courses. *The International Review of Research in Open and Distributed Learning,* 12(7).

[3] Milligan, C., Littlejohn, A. & Margaryan, A., 2013. Patterns of Engagement in Connectivist MOOCs. *MERLOT Journal of Online Learning and Teaching,* 9(2), pp. 149-159.

[4] Sinha, T., 2014. *Supporting MOOC instruction with social network analysis,* Chennai: Department of Computer Science, Vellore Institute of Technology.

References (2/2)

[5] Bergman, Z., 2016. *Analysis of Bath University Massively Open Online Courses using Social Network Analysis,* Bath: Unpublished.

[6] Zhang, J., Skryabin, M. & Song, X., 2016. Understanding the dynamics of MOOC discussion forums with simulation investigation for empirical network analysis (SIENA). *Distance Education,* 05 Sep.pp. 270-286.

[7] Wasserman, S. & Faust, K., 1994. *Social network analysis methods and applications.* 1 ed. Cambridge : Cambridge University Press.

APPENDIX SLIDES

Other Research (1/2)

- Literature which has found the most appropriate significance test and key findings:
  - Ravana et al. (2015) - Information retrieval systems – aim to identify new method for more accurate results. Compared t-test to Wilcoxon rank-sum (Mann Whitney), bootstrapping, permutation. Largely similar results for bootstrap, t-test and permutation. Wilcoxon unreliable for dataset.
  - Biology – comparing microbial community fingerprints – Comparison of p-values for comparing correlations, made power and sample size considerations

Other Research (2/2)

- Other hypotheses to test that were considered:
  - Sinha et al. (2014) – focus on high attrition rates – split by cohort, and centrality – found students who post earlier, continue to contribute and that students with lower centrality are often not discussion facilitators – aim to facilitate development better computer mediated support for students
  - Kellogg et al. (2014) – higher reciprocity for students with common attributes

Network analysis metrics considered

- Centrality – Bergman (2016) did some interesting work relating how central a user is in the network to their success – significant but weak correlation to three success factors
- Betweenness centrality - how an individual acts as a bridge between groups of people – a user with higher rating will have more information pass through them and therefore have the greatest social learning. Representative of the shortest path for which information can travel but does not show how many connections are being made between users – not supporting the hypothesis
- Closeness centrality – how near someone is to the rest of the network - user basis
- Density – which shows number of ties present out of total number of ties – but does not indicate directionality, so one user could be particularly active and distort the trend