Supplementary Information

Appendix for
The Wikipedia Network of Influence Between Painters
&
Community Detection with Metadata in a Network of Influence between Painters

Michael Kitromilidis, Tim S. Evans

Centre for Complexity Science, and Theoretical Physics Group,
Imperial College London, SW7 2AZ, U.K.

10/11/19

This is an appendix of supplementary information for [1] and [2].

S.1 Visualisation of Network

In Fig. S1 is a visualisation of our network of painters based on the edgelist provided in [3].

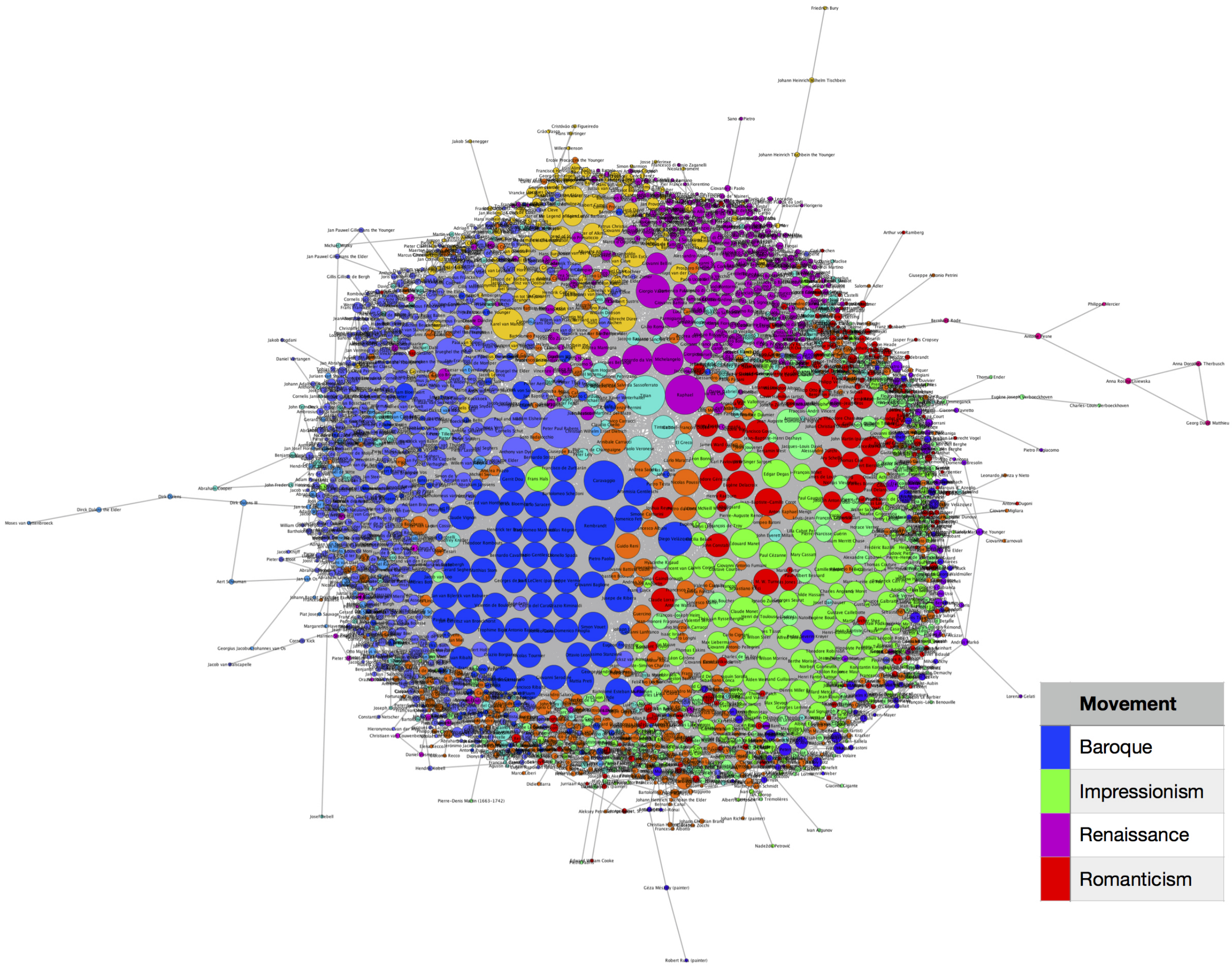

Figure S1: Communities in the painter network; node size corresponds to the degree and colour to the community in which it is placed under the standard implementation of the Louvain method.

S.2 Details about the community partition

The Louvain modularity maximisation method in its standard implementation reveals 14 communities in the painter network (Table S1 displays the 12 significant ones).

Table S1: Significant Louvain communities of painters (number of mentions in brackets).

Tag	Size	Notable Artists	Movements	Locations
#0	171	Turner, Delacroix	Romanticism (75)	French (32), English (31), German (30)
#1	301	Poussin, Caracci	Baroque (212), Rococo (39)	Italian (221)
#2	136	Dürer, van Mander	Northern Renaissance (80)	Flemish (60)
#3	436	Rubens, van Dyck, Breughel	Baroque (353), Mannerism (39)	Dutch (200), Flemish (166)
#4	261	Raphael, Da Vinci, Vasari	High Renaissance (81), Mannerism (63)	Italy (230)
#5	201	Monet, Cézanne, Manet	Impressionism(129), Realism (48)	French (76)
#6	262	David, Ingres	Baroque (61), Neoclassicism (48), Rococo (46)	French (149)
#7	137	Titian, Tintoretto	Baroque (65), Mannerism (22)	Spanish (33)
#8	50	-	Baroque (37)	Dutch (35)
#9	130	Rembrandt, Caravaggio	Baroque (107)	Italian (38), Dutch (27)
#10	52	-	Realism (18), Romanticism (16)	Hungarian (16)
#11	30	-	Realism (20)	Italian (26)

We note that some communities (e.g. 1, 3, 5 and 9) are movement-based, whereas others (e.g. 4, 6) are mainly location-based.

S.3 Metadata for painters

We collect the main artistic movements as identified in the WGA and also what we observe from our analysis in Section A. The tags associated with each node in terms of their location and artistic movement are as follows:

Movement Medieval, Early Renaissance, Northern Renaissance, High Renaissance, Mannerism, Baroque, Rococo, Neoclassicism, Romanticism, Realism, Impressionism

Location American, Austrian, Belgian, Bohemian, Catalan, Danish, Dutch, English, Finnish, Flemish, French, German, Greek, Hungarian, Irish, Italian, Netherlandish, Norwegian, Polish, Portuguese, Russian, Scottish, Spanish, Swedish, Swiss

S.4 Implementation of quality measures: Synthetic Network

To illustrate our implementation of quality measures we consider an artificial network generated by the Stochastic Block Model [9]. We give each node a vector of two hidden attributes $X_i = (x_1^{(i)}, x_2^{(i)})$, and each x_j can take the value of 0 or 1 with equal probability; this means that the possible configurations are $\xi_1 = (0,0)$, $\xi_2 = (0,1)$, $\xi_3 = (1,0)$ and $\xi_4 = (1,1)$. More specifically here $m = 2$, $k_1, k_2 = 2$ and $\nu = 4$.

Two nodes are linked depending on the common attributes they share, i.e. they are linked with probability 1 if they have both attributes matching, with probability 1/2 if one attribute only is matching and are disconnected otherwise.

$$P_{ij} = \begin{pmatrix} 1 & 1/2 & 1/2 & 0 \\ 1/2 & 1 & 0 & 1/2 \\ 1/2 & 0 & 1 & 1/2 \\ 0 & 1/2 & 1/2 & 1 \end{pmatrix} . \quad (1)$$

An optimal partition should uncover four communities in this case; however the standard implementation of modularity only yields two see Fig. S2. In this case the average cluster homogeneity is 0.75, as each community contains two kinds of nodes.

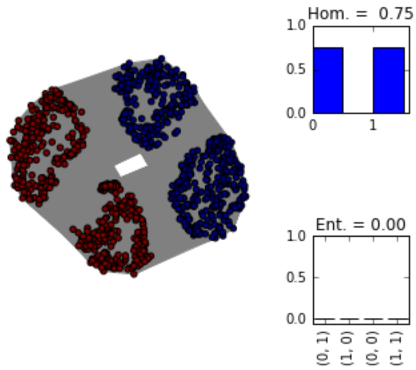

Figure S2: Synthetic network partition with modularity maximisation; two communities detected (underdetection).

This is the opposite scenario from the Karate Club network, as we are *underdetecting* communities; since $h(c) < 1$ while $e(\xi) = 0$ it means that our clusters contain more than one type of node but all the nodes of one type are in a single community. By running Louvain modularity maximisation again in each community (treated as a separate network) we are able to unfold the partition into four communities, as each original community splits into two (Fig. S3). This then leaves us with a perfect result, not surprising here given that this was an artificial model though unlikely to be repeated exactly in a real world context.

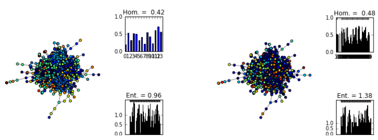

Figure S3: Running modularity maximisation at the communities level uncovers the deeper level clusters and now we get perfect scores for cluster homogeneity, while configuration entropy remains at the optimal level.

S.5 Identifying influential nodes: Mixing parameter

We look at the ratio of the number of links within the community to the number of links outside the community for each node (the *mixing parameter* [10]) given by

$$\mu_{\mathcal{C}}(i) = \frac{k_i^{\text{out}}}{k_i}\,, \qquad k_i^{\text{out}} = \sum_j A_{ij}\left(1 - \delta_{c(i),c(j)}\right)\,, \tag{2}$$

where A_{ij} is the adjacency matrix. Fig. S4 shows the correlation of this measure with standard centrality measures; the correlation is relatively weak, which illustrates that this measure can indeed have a significant contribution in highlighting nodes which the other measures may not identify.

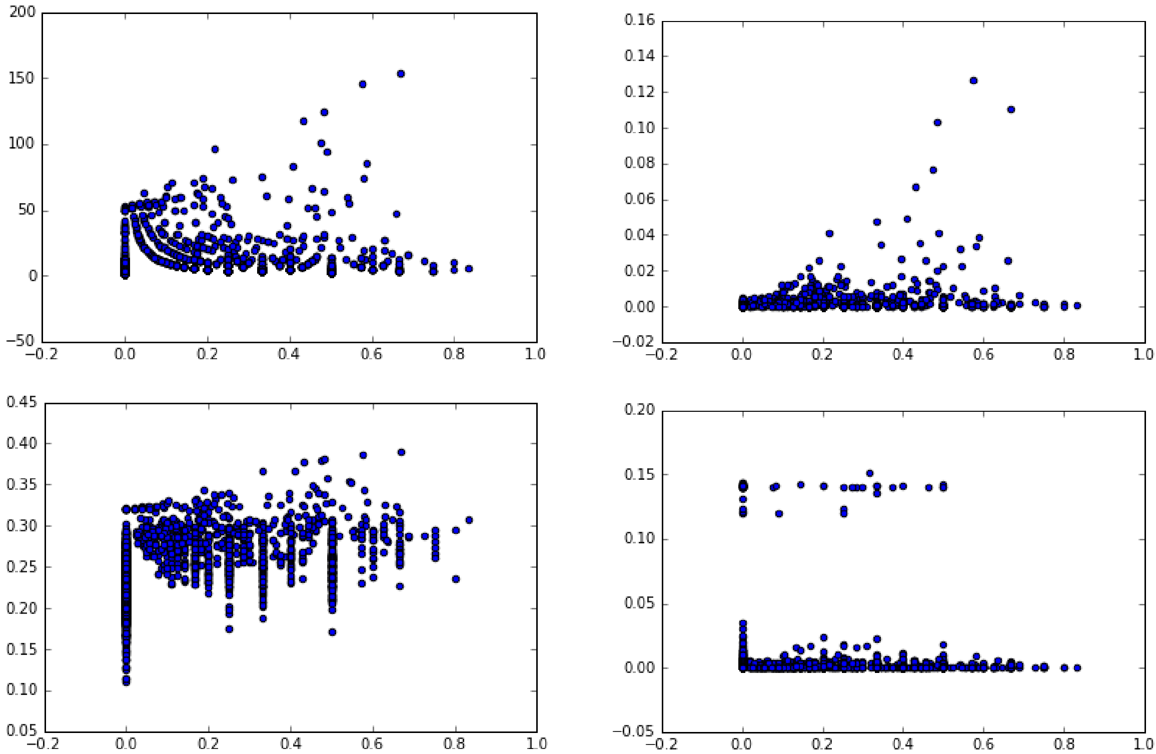

Figure S4: Pearson correlation (horizontal axis) of the mixing parameter $\mu_{\mathcal{C}}(i)$ with centrality measures (vertical axes, clockwise from top left: degree, betweenness, eigenvector and closeness centrality).

S.6 Identifying influential nodes: Community-based betweenness centrality

The correlations between the standard and modified betweenness centrality are quite high for both of our partitions (almost 1). However the ranks of the nodes exhibit smaller correlation values (around 0.88) allowing us to identify certain nodes who score poorly in the standard Betweenness Centrality and better in our Community-Based modification (Fig. S5).

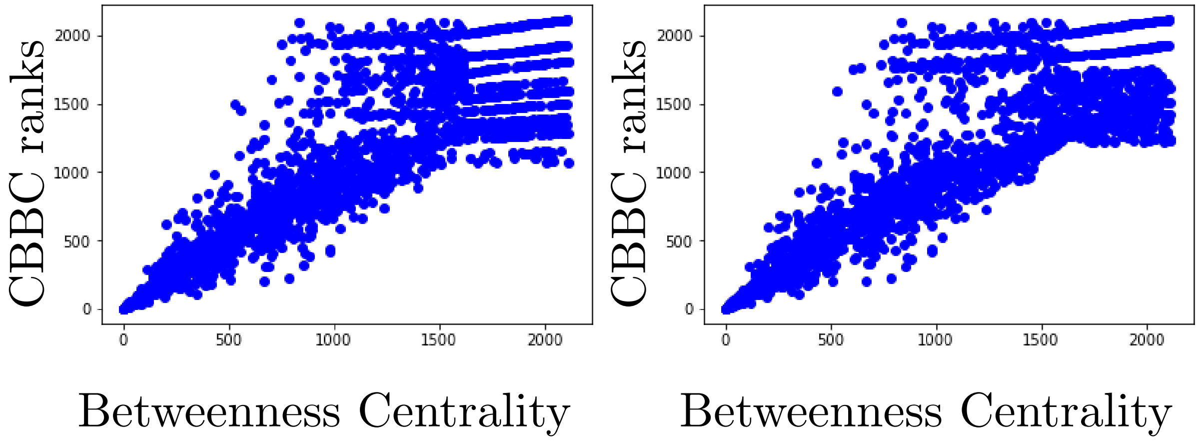

Figure S5: Correlations between standard and community-based betweenness centrality, in the original (left) and fine (right) partitions.

S.7 Examples highlighted by the new measures

To illustrate our methods, we described in [2] painters who are highlighted as being influential (also see [11] and [12]). In Fig. S6 we give examples of paintings which illustrate the influence between the painters highlighted by our methods.

(a1) Chardin

(a2) Cézanne

(b1) Liebermann

(b2) Manet

(c1) Ruisdael

(c2) Constable

Figure S6: Examples of painters highlighted by the community based centrality measures: (a) Example of how Chardin's still life works were highly influential for impressionists, such as Cézanne; note the similarity of laid out objects and the angled knife on the left used to give a depth perspective. (b) Liebermann was influenced by French impressionists such as Manet, whom he had encountered during his stay in Paris; here we see him adopting the more relax posture for his *Portrait of a Seated Lady* from Manet's *Winter Garden*. (c) Ruisdael's *Landscape with windmills* was studied by many artists for his landscape painting techniques, including in this copy by John Constable.

References

[1] M. Kitromilidis, T.S. Evans, "The Wikipedia Network of Influence Between Painters", submitted to Complenet 2020.

[2] M. Kitromilidis, T.S. Evans, "Community Detection with Metadata in a Network of Influence between Painters", submitted to Complenet 2020.

[3] M. Kitromilidis, T.S. Evans, "Painters Network", figshare, fileset, (2018) `http://doi.org/10.6084/m9.figshare.5419216`.

[4] U. Brandes, D. Wagner: visone - Analysis and Visualization of Social Networks. In Michael Jünger and Petra Mutzel (Eds.): Graph Drawing Software, pp. 321-340. Springer-Verlag, 2004.

[5] A.N. Langville, C.D. Meyer, "Who's no.1?: The science of rating and ranking". Princeton University Press, 2012.

[6] R. Bruggemann, B. Münzer, E. Halfon, "An algebraic/graphical tool to compare ecosystems with respect to their pollution — the German river "Elbe" as an example — I: Hasse-diagrams", Chemosphere **28** (1994) 863–872.

[7] T.V. Loach, T.S. Evans, "Ranking Journals Using Altmetrics", figshare.com (2015) `10.6084/m9.figshare.1461693`.

[8] B. Chen, Z. Lin, T.S. Evans, "Analysis of the Wikipedia Network of Mathematicians" arXiv preprint (2019) `arXiv:1902.07622`.

[9] Holland PW, Laskey KB, Leinhardt S. Stochastic blockmodels: First steps. Social Networks. 1983;5(2):109–137.

[10] Fortunato S, Hric D. Community detection in networks: A user guide. Physics Reports. 2016;659:1–44.

[11] Kitromilidis ME. Topics of interdisciplinary applications of complex network theory. PhD Thesis, Imperial College London, 2018. `http://hdl.handle.net/10044/1/62640`.

[12] Kitromilidis ME, Evans TS. Community Detection with Metadata in a Network of Biographies of Western Art Painters. arXiv preprint. 2018; `arXiv:1802.07985`.