Center-Periphery Structure in Communities: Extracellular Vesicles
Supplementary Materials

Eleanor Wedell, Minhyuk Park, Dmitriy Korobskiy, Tandy Warnow, and George Chacko

November 15, 2021

1 Versions and Commands

KC

- Commit ID: 652befd4316ccfac67d689ac3d504afa98431416
- Availability: `https://github.com/chackoge/ERNIE_Plus/tree/master/Illinois/clustering/eleanor/code`
- Command:

```
python KC.py -e <INPUT_NETWORK> -o <OUTPUT_FILE> -k <K_VALUE>
```

IKC

- Commit ID: 652befd4316ccfac67d689ac3d504afa98431416
- Availability: `https://github.com/chackoge/ERNIE_Plus/tree/master/Illinois/clustering/eleanor/code`
- Command:

```
python IKC.py -e <INPUT_NETWORK> -o <OUTPUT_FILE> -k <K_VALUE>
```

Augmentation

- Commit ID: 69fb38759fb2afa84a84ad4859e56b12a8a4c131
- Availability: `https://github.com/chackoge/ERNIE_Plus/commits/master/Illinois/clustering/minhyuk/python_scripts`

- Command:

```
python -m python_scripts.cluster_processing_scripts.assign_unclustered_nodes
--network-file <INPUT_NETWORK> --clustering <CLUSTERING_FILE> \
--cluster-criterion <CRITERION> --config-file <CONFIG_FILE> \
--output-prefix <OUTPUT_PREFIX> --min-degree <<MIN_DEGREE> \
--core-nodes-file <CORE_NODES>
```

Graclus

- Version: 1.2
- Availability: `https://www.cs.utexas.edu/users/dml/Software/graclus.html`
- Command:

```
graclus -l <NUM LOCAL SEARCH> <INPUT NETWORK> <NUM CLUSTERS>
```

Recursive Graclus

- Commit ID: 69fb38759fb2afa84a84ad4859e56b12a8a4c131
- Availability: `https://github.com/chackoge/ERNIE_Plus/commits/master/Illinois/clustering/minhyuk/python_scripts`
- Command:

```
python -m python_scripts.cluster_processing_scripts.recursive_graclus \
--clustering <CLUSTERING_FILE> --network <INPUT_NETWORK> \
--output-prefix <OUTPUT_PREFIX> --k <k> --m <m> --local-search <local_search>
```

Iterative Graclus

- Commit ID: 69fb38759fb2afa84a84ad4859e56b12a8a4c131
- Availability: `https://github.com/chackoge/ERNIE_Plus/blob/master/Illinois/clustering/minhyuk/supplementary_materials_scripts/iterative_graclus.sh`
- Commit ID (for parsing_clusters.py): 83b6eaa210a2b75f3fc2eaf2fadd6ad10eaf365d
- Availability (for parsing_clusters.py): `https://github.com/chackoge/ERNIE_Plus/tree/master/Illinois/clustering/eleanor/code`

Leiden

- Version: 1.1.0
- Availability: `https://github.com/CWTSLeiden/networkanalysis`
- Command:

```
java -cp networkanalysis-1.1.0.jar \
nl.cwts.networkanalysis.run.RunNetworkClustering \
-r <RESOLUTION VALUE> -o <OUTPUT CLUSTERING> <INPUT NETWORK>
```

KMP-Parsing

- Commit ID: 83b6eaa210a2b75f3fc2eaf2fadd6ad10eaf365d
- Availability: `https://github.com/chackoge/ERNIE_Plus/tree/master/Illinois/clustering/eleanor/code`
- Command:

```
python parsing_clusters.py -e <INPUT_NETWORK> -o <OUTPUT_FILE>
-c <CLUSTERING_FILE> -k <K_VALUE> -p <P_VALUE>
```

Strict KMP-Parsing (for Leiden Clusterings)

- Commit ID: 4ead3c7975af71ffed1ad185bcf75a7892d0aca5
- Availability: `https://github.com/chackoge/ERNIE_Plus/tree/master/Illinois/clustering/eleanor/code`
- Command:

```
python parsing_clusters_strict.py -e <INPUT_NETWORK> -o <OUTPUT_FILE>
-c <CLUSTERING_FILE> -k <K_VALUE> -p <P_VALUE}
```

Clustering Statistics

- Commit ID: 51a711b5207dd21ed43ffd273feab16ec07d7e2c
- Availability: `https://github.com/chackoge/ERNIE_Plus/tree/master/Illinois/clustering/eleanor/code`
- Command:

```
python cluster_info.py -e <INPUT_NETWORK> -o <OUTPUT_FILE>
-c <CLUSTERING_FILE> -k <K_VALUE> -p <P_VALUE>
```

SABPQ Expansion

- Commit ID: 020fd08b27323635ef03a26585cb1b351389f950
- Availability: `https://github.com/chackoge/ERNIE_Plus/blob/master/Illinois/clustering/minhyuk/supplementary_materials_scripts/sabpq_expansion.sql`

MDS We used the MDS implementation provided by scikit-learn `https://scikit-learn.org/stable/modules/generated/sklearn.manifold.MDS.html`.

- Commit ID: 3073d02eedfd397b3a2463e5e6c878f13e40494a
- Availability: `https://github.com/chackoge/ERNIE_Plus/blob/master/Illinois/clustering/minhyuk/supplementary_materials_scripts/mds.py`

2 Additional Tables

Table 1: Pairwise intersections of references cited in review articles. The union of these references was used to build the set of marker nodes used in the article. Numbers on the diagonal refer to the total number of references in each article for which we were able to find DOIs.

	busatto	clancy	ghoroghi	he	kallrefs	kallrevs	lananna	leidal	lelay	raposo	schnatz	verdi	vniel
busatto	85	1	9	2	10	0	4	2	0	1	9	2	5
clancy		78	3	4	9	2	7	3	2	1	5	3	17
ghoroghi			161	1	18	2	6	4	1	1	7	8	8
he				61	4	1	5	2	2	2	3	2	3
kallrefs					240	55	30	5	5	3	11	7	17
kallrevs						55	1	2	1	1	2	2	1
lananna							158	7	5	3	3	4	13
leidal								93	1	1	3	4	3
lelay									104	1	0	1	3
raposo										8	1	2	4
schnatz											150	12	11
verdi												112	7
vniel													185

- *busatto:* Busatto et al. (10.1096/fba.2021-00045). The role of extracellular vesicles in the physiological and pathological regulation of the blood–brain barrier.
- *clancy:* Clancy et al. (10.1096/fba.2020-00127). The ins and outs of microvesicles
- *ghoroghi:* Ghoroghi et al. (10.1096/fba.2021-00079). Tumor extracellular vesicles drive metastasis (it's a long way from home)
- *he:* He et al. (10.1096/fba.2021-00040). Plant extracellular vesicles: Trojan horses of cross-kingdom warfare
- *kallrefs:* Kalluri et al. (10.1126/science.aau6977) The biology, function, and biomedical applications of exosomes (all references)
- *kallrevs:* Kalluri et al. (10.1126/science.aau6977) The biology, function, and biomedical applications of exosomes *(review articles only)*
- *lananna:* Lananna (10.1096/fba.2021-00077). Friends and foes: Extracellular vesicles in aging and rejuvenation
- *leidal:* Leidal (10.1096/fba.2020-00138). Emerging roles for the autophagy machinery in extracellular vesicle biogenesis and secretion
- *lelay:* Lelay (10.1096/fba.2020-00147). Adipocyte-derived extracellular vesicles in health and diseases: Nano-packages with vast biological properties.
- *raposo:* Raposo (10.1096/fba.2021-00009). Extracellular vesicles and homeostasis—An emerging field in bioscience research

- *schnatz:* Schnatz (10.1096/fba.2021-00035). Extracellular Vesicles in neural cell interaction and CNS homeostasis.
- *verdi:* Verdi (10.1096/fba.2021-00081). In vivo imaging of EVs in zebrafish: New perspectives from "the waterside"
- *vniel:* van Niel (10.1038/nrm.2017.125). Shedding light on the cell biology of extracellular vesicles

Resolution	Node Coverage (%)	Num Clusters	Num Singletons	Min	Median	Max
0.05	57.0	488,285	6,323,695	2	17	960
0.10	36.2	506,263	9,369,353	2	10	480
0.15	26.2	501,724	10,851,508	2	7	320
0.20	19.1	489,782	11,890,475	2	5	240
0.25	15.6	481,780	12,408,446	2	4	192
0.30	15.7	506,285	12,382,105	2	4	160
0.35	12.0	467,023	12,928,250	2	3	138
0.40	12.1	498,163	12,910,739	2	3	120
0.45	12.2	514,329	12,901,404	2	3	107
0.50	8.7	434,973	13,412,183	2	2	97
0.55	8.6	430,687	13,430,259	2	2	62
0.60	8.5	439,799	13,443,892	2	2	56
0.65	8.5	446,609	13,442,235	2	2	54
0.70	8.2	475,972	13,487,975	2	2	45
0.75	8.2	489,937	13,496,111	2	2	39
0.80	8.1	495,132	13,497,983	2	2	31
0.85	8.1	496,427	13,499,175	2	2	23
0.90	8.1	497,520	13,499,433	2	2	17
0.95	8.1	497,757	13,499,132	2	2	16

Table 2: **Statistics of Leiden Clusterings** Statistics are shown for Leiden clusterings for different settings of the resolution value, under the Constant Potts model (default), and without restriction to *kmp*-valid clusters. Singleton clusters were not included in the statistics.

Resolution	node coverage (%)	# clusters	# Singletons	Min	Median	Max
0.05	3.2	4076	14231603	20	89	960
0.10	2.6	5393	14319177	12	58	480
0.15	2.4	7041	14347520	9	41	320
0.20	2.3	9262	14352224	6	31	240
0.25	2.3	11471	14361736	6	24	192
0.30	1.9	10786	14422904	6	21	160
0.35	2.1	15458	14387295	6	17	138
0.45	1.4	13179	14484578	6	14	107
0.50	1.4	15638	14490772	6	11	97
0.55	1.3	16230	14500339	6	11	62
0.65	1.0	14717	14547985	6	9	54
0.70	0.7	11450	14586986	6	9	45
0.75	0.5	8333	14623377	6	8	39
0.85	0.3	7042	14645600	6	6	23
0.90	0.3	7001	14648741	6	6	17
0.95	0.3	7009	47648697	6	6	16

Table 3: **Statistics of the kmp-valid Leiden clusters** We show the statistics of the Leiden clustering output, restricted to those clusters that are kmp-valid. For this analysis we set $k = 5$ and $p = 2$.

Resolution	node coverage (%)	# clusters	# Singletons	Min	Median	Max
0.05	2.5	2320	14322318	30	137	960
0.10	2.0	2961	14406801	21	88	480
0.15	2.7	3518	14447698	22	63	320
0.20	1.5	3962	14477262	20	50	240
0.25	1.3	4146	14509238	19	41	192
0.30	1.0	3957	14543469	15	35	160
0.35	0.9	3909	14566858	15	30	138
0.45	0.6	3283	14610330	14	24	107
0.50	0.4	2522	14636390	14	22	97
0.55	0.3	2345	14645853	12	20	62
0.65	0.2	1657	14666283	12	17	54
0.70	0.1	1228	14675151	11	16	45
0.75	0.1	770	14683643	11	15	39
0.85	0.0	261	14692039	11	13	23
0.90	0.0	55	14694801	11	11	17
0.95	0.0	57	14694814	11	11	16

Table 4: **Statistics of the *kmp*-valid Leiden clusters** We show the statistics of the Leiden clustering output, restricted to those clusters that are *kmp*-valid. For this analysis we set $k = 10$ and $p = 2$.