

Chandrasekharan et al. (2020) Supplementary Materials

Shreya Chandrasekharan

September 25, 2020

## Article Network Examples

### Cluster #1673

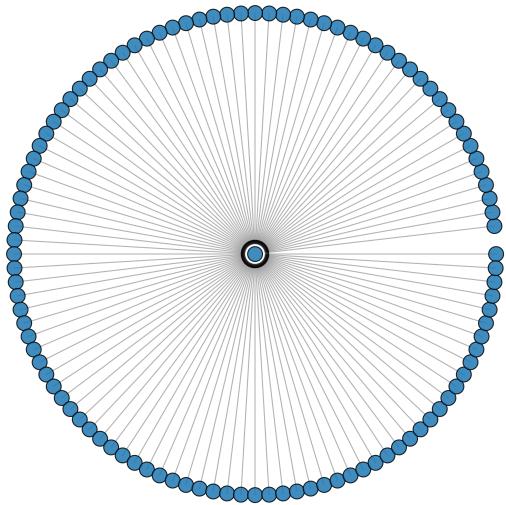


Figure 1: Cluster consisting of 108 publications focused on the comparison of hypersensitivity vasculitis and the Henoch:Schonlein purpura, and is cited by the other 107 publications in the cluster (shown in center).

## Cluster #1301

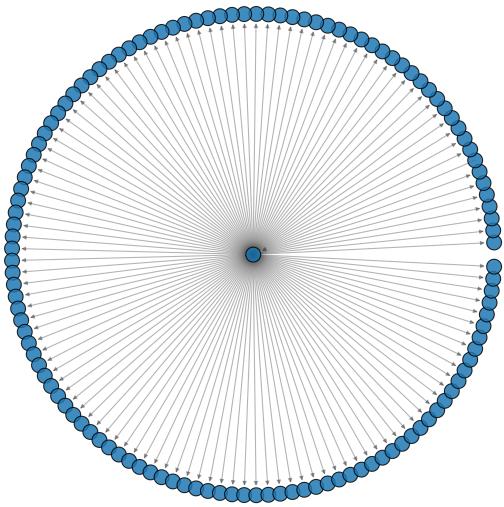


Figure 2: Cluster consisting of 125 publications focused on non-steroidal anti-inflammatory drugs (NSAIDS), and 123 of its articles are cited by a single article from the cluster (shown in center), which is cited, in turn, by another article in the cluster.

## Cluster #8661

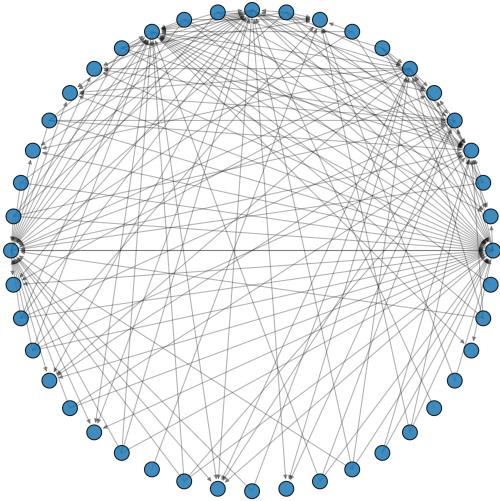


Figure 3: Cluster consisting of 44 publications focused immunoglobulin genes with 19 of them receiving citations from 31 nodes within the cluster.

## Distribution of Author Tiers

Distribution of author tiers at varying thresholds in the pipeline for all three datasets

### imm\_1

Author-tier distribution for size-restricted set (between size 30 and 350)

Threshold: Jaccard Coefficient (JC) of at least 25th Percentile Value (0.20)

Author Tier Distribution	Values
number of authors	1589041
mean number of clusters per author	1.999
minimum number of clusters per author	1
25th percentile number of clusters per author	1
median number of clusters per author	1
75th percentile number of clusters per author	2
maximum number of clusters per author	156
mean tier 1 count per author	0.0572
minimum tier 1 count per author	0
25th percentile tier 1 count per author	0
median tier 1 count per author	0
75th percentile tier 1 count per author	0
maximum tier 1 count per author	37
mean tier 2 count per author	0.175
minimum tier 2 count per author	0
25th percentile tier 2 count per author	0
median tier 2 count per author	0
75th percentile tier 2 count per author	0
maximum tier 2 count per author	50
mean tier 3 count per author	1.766
minimum tier 3 count per author	0
25th percentile tier 3 count per author	1
median tier 3 count per author	1
75th percentile tier 3 count per author	2
maximum tier 3 count per author	139

Threshold: Jaccard Coefficient (JC) of at least 25th Percentile Value (0.20)

Internal-edges to Nodes Ratio (IENR) of at least 99th Percentile Value (1.84)

Author Tier Distribution	Values
number of authors	29469
mean number of clusters per author	1.044
minimum number of clusters per author	1
25th percentile number of clusters per author	1
median number of clusters per author	1
75th percentile number of clusters per author	1
maximum number of clusters per author	4
mean tier 1 count per author	0.065
minimum tier 1 count per author	0
25th percentile tier 1 count per author	0
median tier 1 count per author	0
75th percentile tier 1 count per author	0
maximum tier 1 count per author	2
mean tier 2 count per author	0.057
minimum tier 2 count per author	0
25th percentile tier 2 count per author	0
median tier 2 count per author	0
75th percentile tier 2 count per author	0
maximum tier 2 count per author	3
mean tier 3 count per author	0.922
minimum tier 3 count per author	0
25th percentile tier 3 count per author	1
median tier 3 count per author	1
75th percentile tier 3 count per author	1
maximum tier 3 count per author	4

## imm\_2

Author-tier distribution for size-restricted set (between size 30 and 350)  
 Threshold: Jaccard Coefficient (JC) of at least 25th Percentile Value (0.22)

Author Tier Distribution	Values
number of authors	2363403
mean number of clusters per author	2.339
minimum number of clusters per author	1
25th percentile number of clusters per author	1
median number of clusters per author	1
75th percentile number of clusters per author	2
maximum number of clusters per author	256
mean tier 1 count per author	0.083
minimum tier 1 count per author	0
25th percentile tier 1 count per author	0
median tier 1 count per author	0
75th percentile tier 1 count per author	0
maximum tier 1 count per author	33
mean tier 2 count per author	0.535
minimum tier 2 count per author	0
25th percentile tier 2 count per author	0
median tier 2 count per author	0
75th percentile tier 2 count per author	1
maximum tier 2 count per author	127
mean tier 3 count per author	1.721
minimum tier 3 count per author	0
25th percentile tier 3 count per author	1
median tier 3 count per author	1
75th percentile tier 3 count per author	2
maximum tier 3 count per author	172

Threshold: Jaccard Coefficient (JC) of at least 25th Percentile Value (0.22)  
 Internal-edges to Nodes Ratio (IENR) of at least 99th Percentile Value (1.72)

Author Tier Distribution	Values
number of authors	44978
mean number of clusters per author	1.055
minimum number of clusters per author	1
25th percentile number of clusters per author	1
median number of clusters per author	1
75th percentile number of clusters per author	1
maximum number of clusters per author	6
mean tier 1 count per author	0.096
minimum tier 1 count per author	0
25th percentile tier 1 count per author	0
median tier 1 count per author	0
75th percentile tier 1 count per author	0
maximum tier 1 count per author	3
mean tier 2 count per author	0.152
minimum tier 2 count per author	0
25th percentile tier 2 count per author	0
median tier 2 count per author	0
75th percentile tier 2 count per author	0
maximum tier 2 count per author	5
mean tier 3 count per author	0.806
minimum tier 3 count per author	0
25th percentile tier 3 count per author	1
median tier 3 count per author	1
75th percentile tier 3 count per author	1
maximum tier 3 count per author	6

## eco

Author-tier distribution for size-restricted set (between size 30 and 350)  
 Threshold: Jaccard Coefficient (JC) of at least 25th Percentile Value (0.19)

Author Tier Distribution	Values
number of authors	2837003
mean number of clusters per author	2.124
minimum number of clusters per author	1
25th percentile number of clusters per author	1
median number of clusters per author	1
75th percentile number of clusters per author	2
maximum number of clusters per author	340
mean tier 1 count per author	0.0715
minimum tier 1 count per author	0
25th percentile tier 1 count per author	0
median tier 1 count per author	0
75th percentile tier 1 count per author	0
maximum tier 1 count per author	37
mean tier 2 count per author	0.592
minimum tier 2 count per author	0
25th percentile tier 2 count per author	0
median tier 2 count per author	0
75th percentile tier 2 count per author	1
maximum tier 2 count per author	87
mean tier 3 count per author	1.459
minimum tier 3 count per author	0
25th percentile tier 3 count per author	1
median tier 3 count per author	1
75th percentile tier 3 count per author	1
maximum tier 3 count per author	317

Threshold: Jaccard Coefficient (JC) of at least 25th Percentile Value (0.19)  
Internal-edges to Nodes Ratio (IENR) of at least 99th Percentile Value (2.5)

Author Tier Distribution	Values
number of authors	36236
mean number of clusters per author	1.091
minimum number of clusters per author	1
25th percentile number of clusters per author	1
median number of clusters per author	1
75th percentile number of clusters per author	1
maximum number of clusters per author	15
mean tier 1 count per author	0.121
minimum tier 1 count per author	0
25th percentile tier 1 count per author	0
median tier 1 count per author	0
75th percentile tier 1 count per author	0
maximum tier 1 count per author	9
mean tier 2 count per author	0.172
minimum tier 2 count per author	0
25th percentile tier 2 count per author	0
median tier 2 count per author	0
75th percentile tier 2 count per author	0
maximum tier 2 count per author	5
mean tier 3 count per author	0.797
minimum tier 3 count per author	0
25th percentile tier 3 count per author	0
median tier 3 count per author	1
75th percentile tier 3 count per author	1
maximum tier 3 count per author	9