Using Network Theory to understand the immune response caused by PDH Deficiency

Introduction

Over a twelve-week period, our group investigated Pyruvate Dehydrogenase Complex Deficiency (PDCD), a rare metabolic disease, using Protein-Protein-Interactions Networks in *Saccharomyces cerevisiae.* As a group, we studied different protein pathways to address the lethal build-up of lactic acid in PDCD patients. In this report, we will use different network theory methods to understand interactions within the immune system. PDCD is associated with a strong immune response causing excessive inflammation and fever (Yagasaki et al., 2019). Chemical messengers called cytokines such as IL-6 and TNF- α are produced in excess by dendritic cells and macrophages (Evans et al., 2015). Atallah and colleagues (2020) at the University of Stanford Medicine have developed a directed network called ImmunoGlobe, which records all known intercellular immune interactions. ImmunoGlobe is publicly available and can be exported as a DigGraph object using the Networkx package in Python. Our aim is to use this graph, to study the role of dendritic cells and macrophages as well as their messengers (IL-6 and TNF- α) in the immune system. We first conducted different community-finding algorithms. When dealing with directed networks, a common approach described by Dugué and Perez (2015) is to omit direction and run Louvain which tries to optimise modularity. Leiden algorithm, a modified version of Louvain discussed by Traag and colleagues (2019) can also be implemented on the undirected network. More appropriate solutions for directed graphs will then be discussed. The Surprise algorithm is a different community-finding algorithm that tries to minimise Surprise instead of modularity and takes direction into account (Traag et al, 2015). The Immunoglobe network is multi-layered, and a subset containing only immune cells can be obtained. Kernighan-Lin algorithm finds two clusters from the network by iteratively swapping each pair of nodes to minimise the sum of the weight on all edges cut (Kernighan et al, 1970). By implementing this algorithm in the subset of Immunoglobe, we check if the separation between innate and adaptive immune systems described in the literature (Hoebe et al, 2004) can be detected. The bisection can be used to understand the role of macrophage and dendritic cells compared to other cells. Lastly, we will perform centrality analysis using Eigenvector and Katz centrality to observe the role of the immune components involved in PDCD. Eigenvector centrality is an extended fork of degree centrality which looks at both the number of neighbours a node has but also the centrality value of those neighbours (Newman, 2010). Katz centrality is an alternative centrality measure that performs better on a directed network (Leung et al, 2014).

Approach and Results

Data Obtention and Data Cleaning

Two files were downloaded from the ImmunoGlobe website. One of them contained information on all the edges and the other on all the nodes. Only the file containing information on the edges is needed to obtain our graph The other will be useful at a later stage. A dataframe was obtained using the Pandas package. It contained 2810 rows indicating the number of edges and three columns: the Source Name, Target Name and Edge Effect. There was eight different edge effect and some of them had a positive effect such as “secrete” or “activate”, whereas other had a negative effect on the target node like “Inhibit” or “kill”. To make our analysis meaningful, we decided to keep the three most common positive effects which were “secrete”, “activate” and “recruit”. Edges with a different meaning would complicate the interpretation of the results and community-finding algorithms may perform poorly. Applying the zip function, we add a new column to the dataframe corresponding to the edge. The edge is stored as a tuple of the form (A,B) where A is the source node and has a positive effect on the target node B. Using the *DiGraph* function from the Networkx package, we obtained a directed graph called *G*. We will focus mostly on four nodes: Macrophage, Dendritic cell, TNF- α and IL6. Macrophage and Dendritic cells have an edge with both TNF- α and IL6. The shortest paths between Dendritic /Macrophage and TNF- α /IL6 are both of length of 3.

Reducing the graph to undirected: Louvain and Leiden algorithms

The graph G is directed and thus, the Louvain algorithm is not supported. We can easily obtain a new graph G_1 which is the undirected version of *G* using the *to_undirected* function from Networkx. The different partitions of G_1 are obtained with the *best_partition* function from community_louvain. The output of a single run is nine different communities with sizes varying from 3 to 72 nodes. The smallest partition contains three nodes: T_MAIT which is a cell of the innate immune system and stands for Mucosal associated invariant T cells, Microbial metabolites and Bacterial Metabolites. Microbial and bacterial metabolites are substances produced by microbes and bacteria respectively. MAIT cells are unconventional T cells that recognise metabolites and can rapidly produce an immune response (Hinks and Zhang, 2020). This example demonstrates that the Louvain algorithm can discover a community that is meaningful in terms of immunology even when the graph is reduced to non-directed. When looking at bigger communities, it is hard to find any immunological meaning. We can still observe some interesting properties like the presence of most of the antibodies in the same community. We can change the resolution parameter when running Louvain. The default is 1 and increasing will lead to bigger communities while decreasing will lead to smaller communities. (Lambiotte et al, 2008). Randomness can be an issue when looking at a single run for Louvain, we used the *random_state* parameter to obtain continuously the same output for a single run. Randomness can be addressed by looking at the results after performing multiple iterations. After removing the *random_state* parameter, we ran Louvain 100 times and looked at the community number for the nodes involved with PDCD (Tab. 1).

	Dentritic Cell	TNF- α	IL-6
Macrophage	13	5	29
IL-6	20	12	
TNF- α	1		

Table 1- Results of Louvain algorithm after 100 iterations.
The value in each cell corresponds to the number of times the two nodes were in the same community after 100 runs. The algorithm was performed using the *best_partition* function in the community_louvain module by setting the resolution to 1.

The most common pair of nodes in the same community is Macrophage and IL-6 which is expected when looking at the literature (Fernando et al, 2006). However, dendritic cells and TNF- α are in the same community only once whereas the cytokine has been shown to play a major role in the maturation of the cell (Trevejo et al, 2001).

Louvain algorithm has been shown to yield badly connected communities and can be improved to the Leiden algorithm by incorporating smart and fast local as well as random neighbour moves (Traag et al, 2019). Leiden algorithm can easily be implemented in python using the cdlib package. A new graph G_i had to be obtained from G_1 where the node labels were converted to integers. Using a mapping dictionary as well as its inverse, we can easily translate integers to their respective label. The Leiden algorithm yielded 7 communities of size varying between 2 and 69. Interestingly, the community of three nodes with MAIT cells discussed earlier was also present. Macrophage, Dendritic cell, TNF- α and IL-6 were all in separate communities. Although Leiden algorithm is affected by randomness, the seed of algorithms in the cdlib package is fixed and there is no option to change the parameter. Consequently, running Leiden 100 times leads to the same result for every run and we cannot obtain a summary like Table 1.

Surprise algorithm, an alternative to modularity for directed networks

Traag and colleagues (2015) discussed another community-finding algorithm that relies on surprise instead of modularity. For a directed graph, the Surprise function to minimise is defined as:

$$S = m^{*}D\left(q_{observed} \,||\, q_{\text{expected}}\right)$$

Where m is the number of edges, $q_{observed} = \frac{\sum_c m_c}{m}$ is the fraction of internal edges, $q_{expected}$ is the expected fraction of internal edges and $D(q_{observed} \,||\, q_{expected})$ is the binary Kullback-Leider divergence which measures the distance between the Bernoulli probability distributions $q_{observed}$, 1- $q_{observed}$ and $q_{expected}$, 1- $q_{expected}$ (Kullback and Leider, 1951)

Compared to Louvain, Surprise supports directed networks and is effective at finding many small communities while Louvain underestimates the number of communities. Whenever a partition different from random cannot be found by Surprise, it will not likely be found by modularity (Traag et al, 2015). In our run, the Surprise algorithm found 100 communities. The community of MAIT cells was also detected using surprise instead of modularity. As for Leiden, all nodes involved in the immune response to PDCD were in different communities. The results did not change across multiple runs as the seed could not be modified. Compared to Leiden and Louvain, Surprise is the algorithm with the maximum group degree centrality for a single community. However, Surprise also has the lowest mean group degree centrality because it finds many communities of small size.

Detecting innate and adaptive systems using Kernighan-Lin bisection

Recall that the ImmunoGlobe network is multilayered. For example, IgE is an antibody whereas IL-2 is a cytokine. The file describing the nature of each node in the immune system is stored as *df_nodes*. We aim to obtain a graph describing the interaction between every cell of the immune system. By filtering through all edges of the whole graph G, we obtained a new graph G_{cell} that contains edges where both the source and the target are immune cells. G_{cell} contains 52 edges and 34 nodes, and we can apply Kernighan-Lin bisection once it is converted to undirected. For an unweighted graph, each edge can be set as having a weight of 1 and the algorithm minimises the number of crossing edges. Given that the algorithm focuses on crossing edges, it is clear why it does not support directed networks. After converting the network to undirected, we hope to obtain two clusters of cells: one corresponding to the adaptive immune system and the other to the innate immune system. The dataset contains no information to determine if the cell belongs to the adaptive or innate immune system. We had to come up with an innovative approach to score the performance of the model. Cells of the adaptive immune system are all subtypes of T or B cells. By filtering through all the cells, we obtain a list containing all the T and B cells (Alberts and al, 2004). We can then score the model by looking at the proportion of T/B cells in each bisection. The algorithm fits a cluster with a proportion of 70% of adaptive cells and another with 30% of adaptive cells, thus it is able to successfully distinguish innate and adaptive immune cells (Fig 1). The algorithm fails for both macrophage and dendritic cell which are in the first bisection corresponding to the adaptive immune system (Fig 1.a).

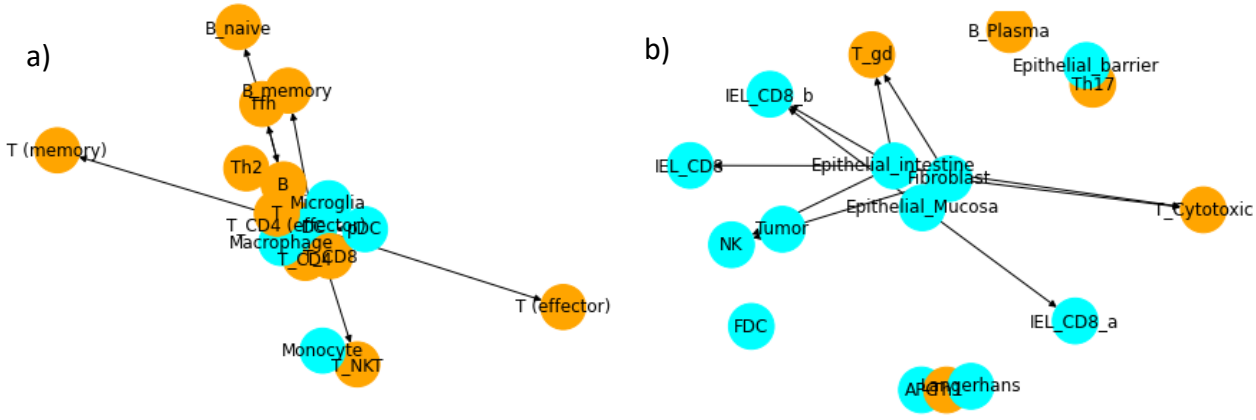


Figure 1 – Bisections obtained by Kernighan-Lin algorithm on immune system cells.
a) Fitted adaptive system: First bisection which is composed of 70% of adaptive cells.
b) Fitted innate system: Second bisection which is composed of 70% of adaptive cells. Adaptive cells are represented in orange and innate cells in cyan.

One limitation of the graphs above is that it only contains cells that are neighbours, that is have a direct edge in *G*. Immune cells usually use messengers called cytokines to communicate. Thus, it is possible for two cells to have a strong interaction in the immune system even if they are not direct neighbours in *G*. Using the *combinations* function from the itertools package, we looked at all possible interactions between pairs of cells. If two cells had a path length less or equal to three, we

stored the edge between the two cells. We formed a new graph G_{long} which contained all edges obtained by the itertools package. It is important to understand that G_{long} is a simplified version of the immune system. Two cells A and B could have a direct edge whereas a middle node corresponding to a messenger should be present. We proceed to run Kernighan-Lin bisection on this simplified graph. Surprisingly, the algorithm performed poorly compared to G_{cell} and one partition had 44% of adaptive cells versus 22% for the other (Supplementary Fig 2). The dataset contains 25% of adaptive immune cells and thus Kernighan-Lin algorithm unsuccessfully distinguishes the adaptive and innate systems on G_{long}.

Centrality analysis focusing on the immune response to PDCD

Dendritic cells and macrophages are overstimulated and release TNF- α and IL-6 during the immune response to PDCD. These components play a capital role in the immune system, and we can use centrality measures to determine if this role is transcribed in the network. Using the *eigenvector_centrality* function from the Networkx package, Macrophage is the node with the highest centrality (0.28) and dendritic cell sits in the third position (0.22). TNF- α and IL-6 which are both cytokines released in PDCD follow dendritic cell closely (0.19) and are fourth and fifth respectively out of the 323 nodes in *G*. Each component stimulated by PDCD plays a central role in the immune system as outlined by their rank and the centrality distribution on the whole graph G (Fig 2a). By overstimulating the major components of the network, we can expect PDCD to cause a massive disruption in the immune system. In theory, eigenvector centrality works for directed networks. However, directed networks usually have an asymmetric adjacency matrix leading to two sets of eigenvectors: left and right eigenvectors. It is not always clear which eigenvector should be used (Neman, 2010). An alternative is Katz centrality. For a directed network, Katz centrality for each node can either be calculated using broadcast centrality where the adjacency matrix A is unchanged or receive centrality which uses the transpose of A denoted A^T (Zhan et al, 2017). Broadcast centrality is the most common one and is defined by Katz (1953) as:

$$\boldsymbol{x} = (I - \alpha A)^{-1}$$

where I is the identity matrix and α is a free parameter that governs the balance between the eigenvector term and constant term for each node. The *katz_centrality* function from the Networkx package use broadcast centrality defined for each node I as

$$x_i = \alpha \sum_j A_{ij}\, x_j + \beta$$

As a default α is set to 0.1 and β to 1. The value of Beta affects the absolute magnitude of centrality which we are not interested in. Applying the function to the directed graph *G*, macrophage and dendritic cell have a high centrality of 0.26 and 0.22 respectively and are ranked first and third across all nodes. The rank of TNF- α and IL-6 has changed compared to eigenvector centrality and they sit at the 6^{th} and 7^{th} place respectively with a centrality of 0.16. We see a very similar distribution compared to Eigenvector centrality (Fig 2b).

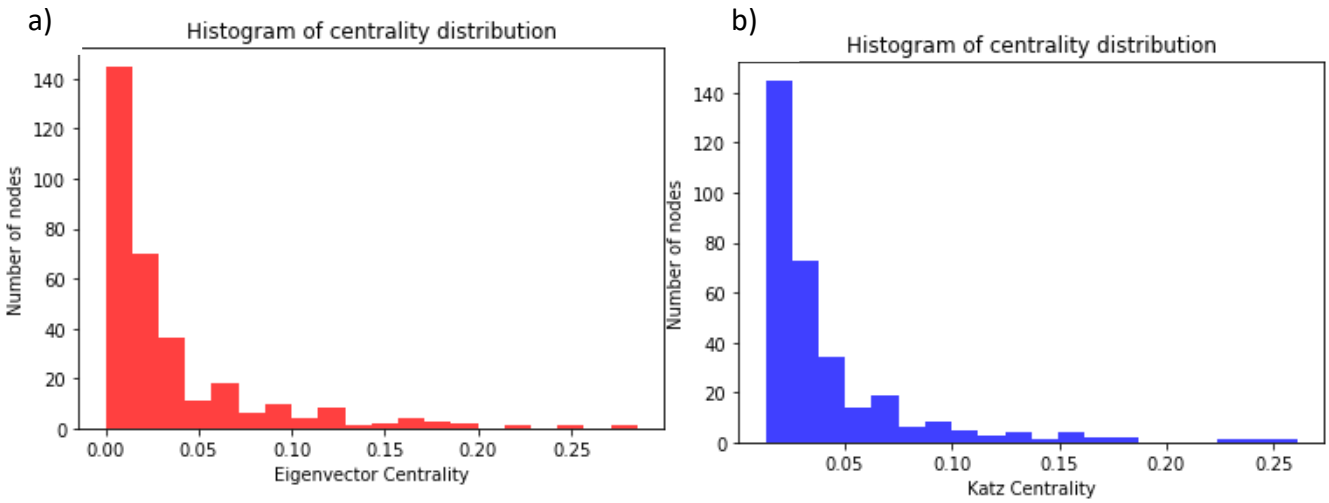


Figure 2 – Centrality distribution of the directed immune system network
a) Histogram of Eigenvector centrality on all nodes of the directed ImmunoGlobe network (Macrophage = 0.28, Dendritic cell = 0.22, TNF- α = IL-6 = 0.19)
b) Histogram of Katz centrality on all nodes of the directed ImmunoGlobe network (Macrophage = 0.26, Dendritic cell = 0.22, TNF- α = IL-6 = 0.16)

When plotting the Katz centrality versus Eigenvector centrality for each node on a scatter plot, we can see that all points lie close to the line y = x indicating that Eigenvector and Katz centrality have nearly equal values for all nodes (Fig 3).

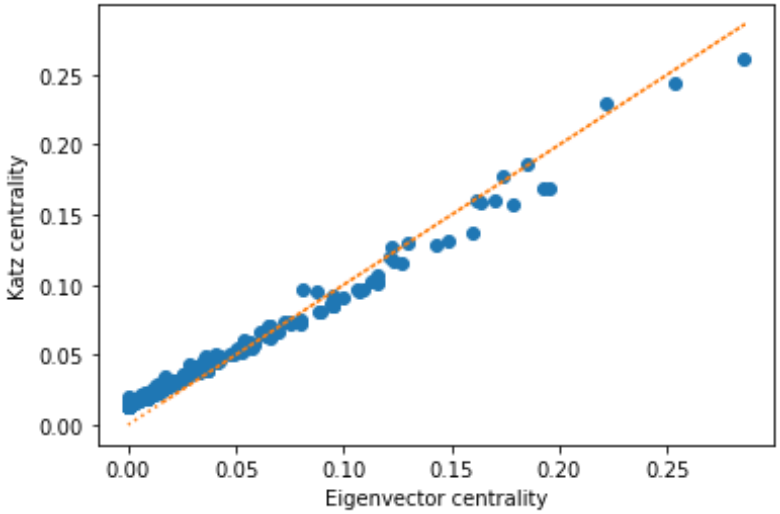


Figure 3 – Scatter plot of Katz against Eigenvector centrality for each component of the immune system.
The orange-dotted line is the identity line of y = x. Katz and Eigenvector centrality have similar results using the networks function as expected when $\alpha_{max}^{-1} \approx 0.1$

Using the *adjacency_spectrum* function on *G*, we can find the value α_{max} to obtain the largest eigenvalue to the adjacency matrix. The algorithm fails to converge for α_{max} even when setting the maximum iterations to 100 000. Setting the maximum iterations to 10 000, the algorithm fails to converge for all values of α greater than 0.1. Approaching α to 0 will make the centrality constant

at each node which is not in our interest. Note that α_{max} is 7.8 and its inverse is 0.12 which is very close to the 0.1, the default value of α
in the *katz_centrality* function. When α is set to 1/ α_{max} Katz and Eigenvector centrality are equal which explains the linear pattern on the scatter plot (Fig 3).

Discussion

Our first approach to detect communities was to reduce the graph to undirected leading to the loss of important information regarding the ImmunoGlobe network. Nevertheless, the Louvain algorithm was able to find some communities that were meaningful when comparing the results to the literature in the field of immunology. When the resolutions in the Louvain algorithm is let to the default one, the Leiden algorithm found the same number of communities and performed very similarly as indicated by the group degree centrality values.

The extension of modularity to directed graphs has been mentioned previously in the literature (Fortunato, 2010). Similarly to Surprise, the probability that an edge is directed in either direction depends on the degree of the target and source nodes. This approach has however been criticised and a different definition of modularity based on diffusion and inspired by PageRank algorithm has been discussed (Dugué and Perez, 2015). Unfortunately, none of the directed versions of Louvain is present in Networks or cdlib packages and we could not implement it on the graph G in Python.
The surprise algorithm was chosen as an alternative to Louvain and Leiden and was implemented on the directed graph *G*. As mentioned by Traag and colleagues (2015), Surprise algorithm overestimated the number of communities with 55 out of the 100 communities with less than three nodes. A parameter similar to *resolution* would be helpful to influence the size of the communities detected by the algorithm. The surprise algorithm supports directed networks and the definition of the function to minimise is the same for directed and undirected graphs (Marchese et al, 2022). The ability to control the seed and thus randomness would have been helpful to compare Leiden and Surprise algorithms to Louvain. Oslom is an alternative detection algorithm that can perform better than Louvain on small-directed networks (Dugué and Perez, 2015). The strength of Oslom is that it can take direction into account when fitting the best partition on the directed graph G (Lancichinetti et al, 2011). Oslom is available with the pyslom package which supports *DiGraph* objects from Networks. However, issues were encountered when installing the package in *Conda* and thus the algorithm was not implemented on *G*.

A possible limitation of the community finding algorithm discussed above is that the graph G has multiple layers corresponding to the type of each node. Focusing only on immune cells, Kernighan-Lin bisection was able to effectively distinguish the adaptive from the innate immune system cells when looking at direct neighbours. The direction from the network had to be removed but this did not affect the model performance. Kernighan-Lin algorithm does not guarantee each partition to be a connected network and indeed it is not the case (Fig 1). However, this is not an issue as similar cells of the immune system such as different B cells do not often communicate with each other. Macrophage and dendritic cells are present in the wrong partition. This can be explained as they serve as antigen-presenting cells and thus have a very strong interaction with B and T cells (Kasheem et al, 2017). When a new edge was added between cells having a path of length three in the whole graph, the Kernighan-Lin algorithm performed poorly to distinguish adaptive and innate cells. When thinking about the interaction between adaptive and innate systems, this can be expected. Indeed, cells of the innate and adaptive immune system communicate with each other using messenger and

thus many edges between innate and adaptive cells were added to the new graph causing Kernighan-Lin to miss the appropriate cut separating each system.

Eigenvector and Katz Centrality both showed similar results to determine the centrality of each node on the directed graph G. PageRank was also used and found different values, but the rank of each cell remained very similar (Supplementary Figure 1). Centrality analysis was useful to analyse the role of the components playing an important role in the immune response to PDCD. PDCD overstimulates cells and cytokines that are central to the immune system leading to numerous symptoms.

Conclusion

As a group, we focused on a different protein pathway called the lipoic acid pathway to limit the effect of PDCD. This individual report presents an alternative looking at the effect of PDCD on the immune system. Although Louvain omit direction, it was able to find communities that had an immunological significance. Using Leiden, and Surprise algorithms we found that macrophage, dendritic cell, TNF- α and IL6 were all in a different community. Kernighan-Lin's algorithm was able to separate adaptive from innate cells but was incorrect for macrophage and dendritic cells. Eigenvector and Katz centrality analysis showed that the four nodes involved in the immune response to PDCD had a major role in the immune system. A possible treatment could be to block the release of cytokines such as TNF- α and IL-6 and thus limiting inflammation and fever due to PDCD. Drugs blocking both TNF- α and IL-6 are already undergoing human clinical trials for patients with diabetes (Genovese et al, 2020) and could be part of a treatment to reduce the symptoms of PDCD. This would need to be used in combination with other treatments as it only avoids the symptoms but does not address the main problem of PDCD which is a build-up of lactic acid in the body. Further investigation from immunology students is required.

Resources

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P., 2002. Lymphocytes and the cellular basis of adaptive immunity. In Molecular Biology of the Cell. 4th edition. Garland Science.

Atallah, M.B., Tandon, V., Hiam, K.J., Boyce, H., Hori, M., Atallah, W., Spitzer, M.H., Engleman, E. and Mallick, P., 2020. ImmunoGlobe: enabling systems immunology with a manually curated intercellular immune interaction network. BMC bioinformatics, 21(1), pp.1-18

Evans, S.S., Repasky, E.A. and Fisher, D.T., 2015. Fever and the thermal regulation of immunity: the immune system feels the heat. Nature Reviews Immunology, 15(6), pp.335-349.

Fernando MR, Reyes JL, Iannuzzi J, Leung G, McKay DM. The pro-inflammatory cytokine, interleukin-6, enhances the polarization of alternatively activated macrophages. PLoS One. 2014 Apr 15;9(4):e94188. doi: 10.1371/journal.pone.0094188. PMID: 24736635; PMCID: PMC3988054.

Fortunato, S., 2010. Community detection in graphs. Physics reports, 486(3-5), pp.75-174.

Genovese, M.C., Burmester, G.R., Hagino, O., Thangavelu, K., Iglesias-Rodriguez, M., John, G.S., González-Gay, M.A., Mandrup-Poulsen, T. and Fleischmann, R., 2020. Interleukin-6 receptor blockade or TNFα inhibition for reducing glycaemia in patients with RA and diabetes: post hoc analyses of three randomised, controlled trials. Arthritis research & therapy, 22(1), pp.1-12.

Hinks, T.S. and Zhang, X.W., 2020. MAIT cell activation and functions. Frontiers in immunology, 11, p.1014.

Hoebe, K., Janssen, E. and Beutler, B., 2004. The interface between innate and adaptive immunity. Nature immunology, 5(10), pp.971-974.

Holtan, S.G., Dronca, R.S., Nevala, W.K., Porrata, L.F., Mansfield, A.S., Block, M.S., Leontovich, A.A., Grotz, T.E., Turner, J.D., Frisch, H.P. and Markovic, S.N., 2011. The dynamic human immune response to cancer: it might just be rocket science. Immunotherapy, 3(9), pp.1021-1024.

Kashem, S.W., Haniffa, M. and Kaplan, D.H., 2017. Antigen-presenting cells in the skin. *Annu Rev Immunol*, *35*(1), pp.469-499.

Katz, L., 1953. A new status index derived from sociometric analysis. Psychometrika, 18(1), pp.39-43.

Lambiotte, R., Delvenne, J.C. and Barahona, M., 2008. Laplacian dynamics and multiscale modular structure in networks. arXiv preprint arXiv:0812.1770.

Leung CK-S, MacKinnon RK, Jiang F. Reducing the search space for big data mining for interesting patterns from uncertain data. In: IEEE international congress on big data (BigData Congress); 2014.

Newman, M., 2010. Networks: An Introduction. *Oxford Academic*

Marchese, E., Caldarelli, G. and Squartini, T., 2022. Detecting mesoscale structures by surprise. *Communications Physics*, *5*(1), pp.1-16.

Nicolas Dugué, Anthony Perez. Directed Louvain: maximizing modularity in directed networks. [Research Report] Université d'Orléans. 2015. ffhal-01231784f

Traag, V.A., Waltman, L. and Van Eck, N.J., 2019. From Louvain to Leiden: guaranteeing well-connected communities. Scientific reports, 9(1), pp.1-12.

Traag, V.A., Aldecoa, R. and Delvenne, J.C., 2015. Detecting communities using asymptotical surprise. Physical Review E, 92(2), p.022816.

Trevejo JM, Marino MW, Philpott N, Josien R, Richards EC, Elkon KB, Falck-Pedersen E. TNF-alpha -dependent maturation of local dendritic cells is critical for activating the adaptive immune response to virus infection. Proc Natl Acad Sci U S A. 2001 Oct 9;98(21):12162-7. doi: 10.1073/pnas.211423598. Epub 2001 Oct 2. PMID: 11593031; PMCID: PMC59785.

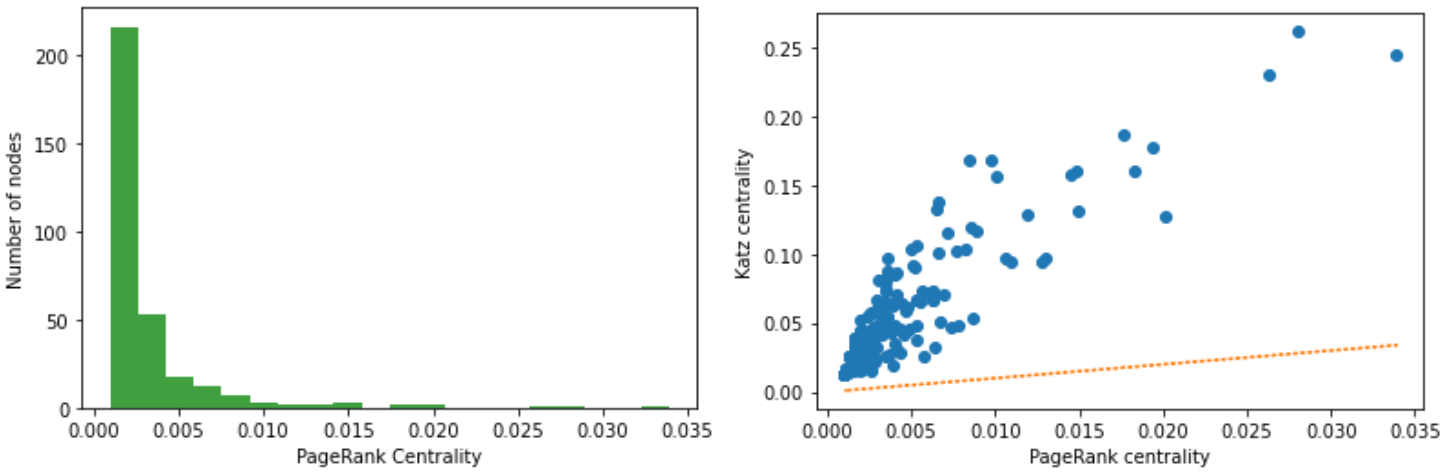
Yagasaki, H., Ohyama, T., Narusawa, H. and Nakane, T., 2019. Hypothermic reaction after infection in an infant with pyruvate dehydrogenase complex deficiency. Pediatrics & Neonatology, 60(4), pp.475-476.

Zhan, J., Gurung, S. and Parsa, S.P.K., 2017. Identification of top-k nodes in large networks using katz centrality. Journal of Big Data, 4(1), pp.1-19.

Appendix

a)

b)



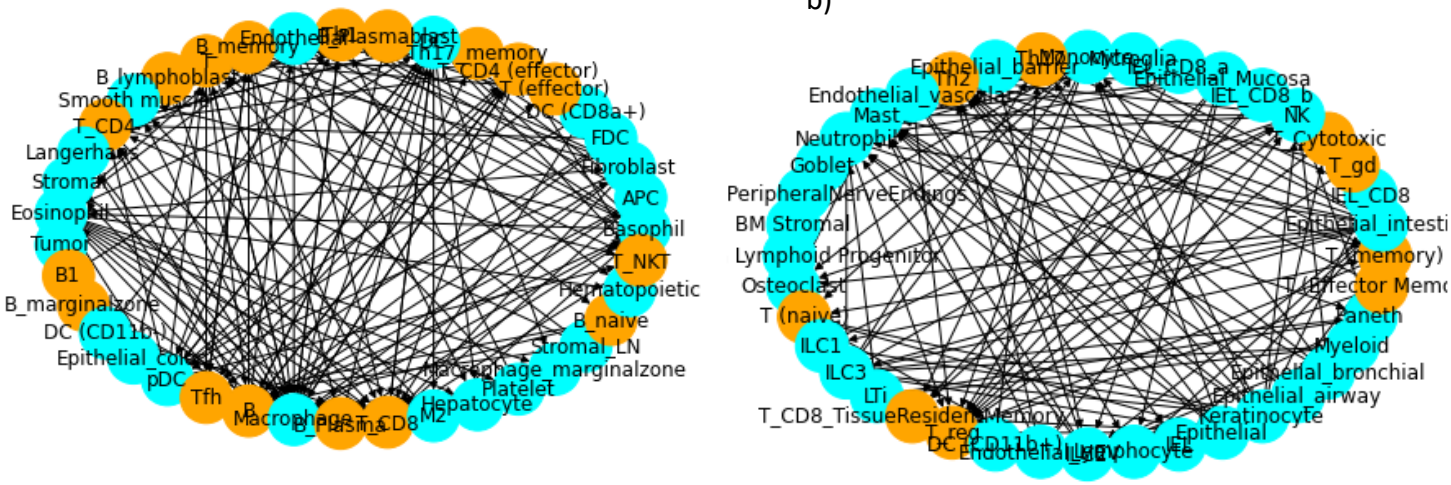
Supplementary Figure 1 – PageRank Centrality distribution

a) Histogram of PageRank Centrality distribution on the directed ImmunoGlobe network (Macrophage = 0.028, Dendritic cell = 0.026, IL-6 = TNF- α = 0.009)

b) Scatter plot of Katz against Eigenvector centrality for each component of the immune system. The orange-dotted line is the identity line of y = x. PageRank and Katz centrality have very different values but the rank of the nodes across the network is conserved

a)

b)



Supplementary Figure 2 - Kernighan-Lin bisection using a simplified network of immune cells

a) Fitted adaptive system: First bisection which is composed of 44% of adaptive cells.

b) Fitted innate system: Second bisection which is composed of 22% of adaptive cells. Adaptive cells are represented in orange and innate cells in cyan.