

An Organic Visual Metaphor for Public Understanding of Conditional Co-occurrences

Keshav Dasu*

Takanori Fujiwara†

Kwan-Liu Ma‡

University of California, Davis

Abstract— Decisions made by domain experts, such as in healthcare and market research, are influenced by the conditional co-occurrence of different events. Learning about conditional co-occurrence is also beneficial for non-experts—the general public. By understanding the co-occurrences of diseases, it is easier to understand which diseases individuals are susceptible to. However, co-occurrence data is often complex. In order for a public understanding of conditional co-occurrence, there needs to be a simpler form to convey such complex information. We introduce an organic visual metaphor. Our model can provide a summary of the conditional co-occurrences within a large set of items and is accessible to the public with its organic shape. We develop a prototype application offering not only an overview for users to gain insights on how co-occurrence patterns evolve based on user-defined criteria (e.g., how do sex and age affect likelihood), but also functionality to explore the hierarchical data in-depth. We conducted two case studies with this prototype to demonstrate the effectiveness of our design.

Index Terms—Visual metaphor, conditional probability, conditional co-occurrence, egocentric visualization

1 INTRODUCTION

In many domains, conditional co-occurrence—the likelihood of other events given that one event has occurred—is used for understanding associated events and making decisions. For example, market researchers are interested in customer purchase patterns (e.g., the likelihood of a customer purchasing product A, given that he/she has purchased product B). The retail industry uses co-occurrences of purchases between different product categories to manage inventory and pricing decisions [2, 3, 32]. Also, microbiologists have an interest in observing co-occurrence of various plankton and bacteria populations to understand the coexistence within biological communities [36]. In health care, conditional co-occurrence of diseases/disorders (or comorbidities [11]) shows the presence of one or more additional diseases/disorders co-occurring with the primary disease/disorder. Therefore, co-occurrence relates to the progression of diseases and mortality risks [9]. The understanding of co-occurrence drives many critical clinical decisions, such as choice of drug therapies, surgical procedures, etc [20, 38].

However, learning about co-occurrence is not limited to domain experts. For example, in healthcare, the public, including patients, could understand the potential consequences of some diseases on their quality of life by understanding the co-occurrence of diseases. It is reasonable to assume a person who is diabetic will be interested in knowing they have a high likelihood of becoming visually impaired. By understanding co-occurrence of diseases, we can make better choices to improve our personal health [19]. As a result, this improves our health literacy and leads directly to improved health outcomes [18]. However, the collected co-occurrence data by researchers is often complex [6, 8, 34]. Thus, for non-domain experts, or the general public, it is difficult to interpret this data and build their own useful insights. Therefore, we need to provide a proper method to share co-occurrence data with the public.

In an effort to improve public understanding of co-occurrence, we introduce an organic visual metaphor, which is designed for not only providing a summary of complicated co-occurrence data but also accessibility to the public [14, 35]. This metaphor is an extended ver-

sion from our previous work for PacificVis Visual Storytelling Contest [7]. In this paper, we describe our metaphor design and the algorithm for generating the visualization. We showcase our metaphor model with an interactive application, through which the user can review co-occurrence of diseases and influences of demographic factors (e.g., ages and sex) on the co-occurrence. We also present two examples that illustrate the utility of our model.

2 RELATED WORK

Conditional co-occurrences are analyzed and interpreted primarily through the use of tables or simple charts (e.g., histograms, simple networks, line charts, pie, and Venn-diagrams). Domain experts in health care typically interpret comorbidities in such a way [6, 8, 34]. It is easier and quicker for domain experts (e.g., doctors) to create and use these formats to understand co-occurrence due to its familiarity. However, when co-occurrence data is large, the results produced with these formats can easily overwhelm both non-domain and even domain experts.

Several systems have been built to help domain experts analyze large conditional-co-occurrence networks [17, 25, 37, 39]. Moni et al. [25] developed CytoCom for visualizing the entire human disease comorbidity network. Huang et al. [17] adapted a Sankey diagram to depict polymorbidity associated with chronic kidney disease. Yang et al.'s [37] system investigates sleep apnea comorbidities using multiple representations: a parallel coordinates, a bubble chart, and a detailed view of comorbidities. Zhou and Mueller [39] developed a framework to explore comorbidities between diseases by overlaying a heatmap over a comorbidity network. Compared with these systems, our work focuses on providing a method for the public to understand complex co-occurrence data by employing a visual metaphor.

3 DESIGN CRITERIA

An important consideration when designing a visualization tool for non-experts to understand and explore complex data is aesthetics. If certain aesthetics are well-embedded in a visualization, it is easier for the viewer to process the information presented. For example, in graph drawing, Purchase [28] found that minimizing edge crossing has a strong effect on human understanding. Cawthon and Moore [1] performed a study to determine the effect of aesthetic on usability in data visualization. They found that people are less likely to abandon a task when they perceive the visualization to have high quality. This implies that people are likely to be more patient when visuals are aesthetically appealing.

*e-mail: kdasu@ucdavis.edu

†e-mail: tfujiwara@ucdavis.edu

‡e-mail: ma@cs.ucdavis.edu

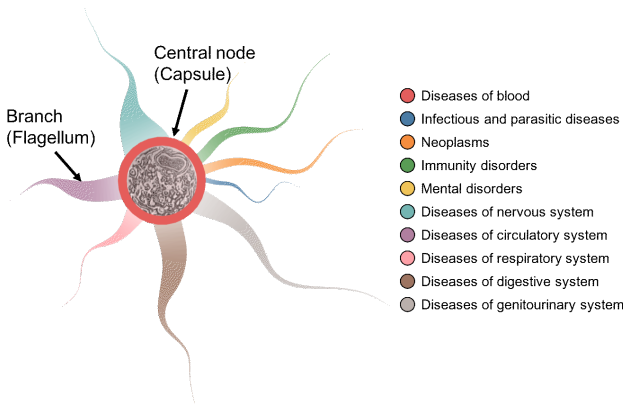


Fig. 1: A bacterium metaphor generated with our model. The central node and branches mirror the bacterium's capsule and flagella.

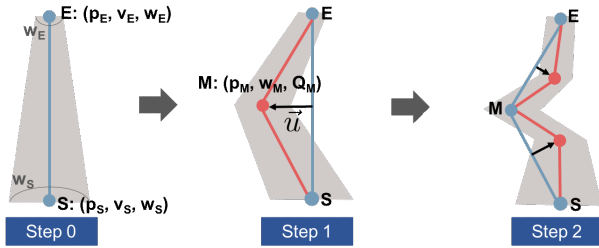


Fig. 2: Recursive steps for generating a flagellum with L-system. Gray backgrounds show the generated flagellum. Blue and red lines are segments before and after applied L-system respectively.

Visual metaphors provide a desirable aesthetic to the viewers by employing a more familiar representation. For example, Wang et al. [35] studied how users react to different personal visualization designs. They found that illustrative designs are the most well-balanced for relaying information and motivating data explorations. Sallaberry et al. [33] employed a biological metaphor to effectively illustrate multiple relationships in a hierarchical dataset. Fung et al. [15] developed an egocentric visualization based a tree metaphor. In a later design study [14], they compared their organic visual metaphor against a set of visualizations and found that their visually attractive approach was the best for at-a-glance characterization of the data. While evaluating the value of the visual metaphors, Risch [29] argued that the users understand visualizations through image schemas—structural patterns established in their childhood. By employing familiar visual metaphors, we enable the user to interpret the data in terms of existing internal schemas. That is, visual metaphors help viewers translate abstract or complex concepts into a form that is more understandable.

In order to make the public understand conditional co-occurrence data more easily, we consider two major design criteria in our visualization: 1) effectively summarizing large, complex data and 2) utilizing a visual metaphor. First, summarizing data effectively is important to avoid overwhelming viewers. To accomplish this, we utilize the hierarchical nature of many commonly occurring data sets. Hierarchical data permits us to bundle elements that belong to the same group [16]. For example, in market research, conditional co-occurrences of purchase of carrots or onions with purchase of cheese or yogurts can be summarized as a co-occurrence of dairy product purchase with purchase of vegetables. Also, in order to clarify the direction of the conditional relationship of co-occurrence, we use an egocentric visualization. Second, as described above, visualizing data with familiar representations for the viewers, including non-experts, can help better understand and motivate the exploration of complex information.

4 VISUALIZATION DESIGN

Based on the design criteria, we derive our organic metaphor model and algorithm to generate the visualization.

4.1 Organic Metaphor Model

Before visualizing conditional co-occurrence, we need to obtain a metric. Here, as one concrete example, we use an association measure q which was used in previous studies in the biomedical science field [10, 24]. However, we can use a different metric based on the needs of the application. q is calculated as follows:

$$q = \frac{P(X|Y)}{P(X)} = \frac{P(X \cap Y)}{P(X)P(Y)} = \frac{N_{X \cap Y}/N}{(N_X/N) \cdot (N_Y/N)} \quad (0 \leq q \leq \infty) \quad (1)$$

where X and Y are two different items (e.g., diseases), $P(\cdot)$ denotes the probability, N is the population size of a given subject group (e.g., patients), N_X and N_Y are numbers of individuals who have at least one occurrence of item X and Y over a given period (e.g., in five years) respectively, and $N_{X \cap Y}$ is a number of individuals who have had occurrences of both X and Y in the same period. The value q measures how strongly the occurrence of X is impacted by the occurrence of Y . When $q < 1$, $q = 1$, or $q > 1$, it means there either is a negative, neutral, or positive association between X and Y , respectively. To specify, when $q > 1$, X and Y are more strongly associated as q increases. For example, when q is a large value, the occurrence of X is significantly increased given that Y has occurred. We consider only the positive associations (i.e., $q > 1$) and use $\log(q)$ as a value v of the co-occurrence metric for the purpose of this paper.

For the visualization, we developed an organic metaphor model. Fig. 1 shows an example of visualizing the co-occurrence of diseases using our model. The visualization not only summarizes the necessary information in an egocentric way, but also motivates users to explore the data [14, 35]. In order to reflect the negative connotation about diseases associations, our design employs a bacterium metaphor. The central node and branches correspond to the bacterium's capsule and flagella. Our organic metaphor visualizes places the focal category (e.g., diseases of circulatory system) in the capsule, while other related categories are placed in the branches or flagella. The colors used for the central node and branches represent the respective category of each item. To further refine our model, various statistical measures related to v in the co-occurrence metric are also visually encoded: 1) the size of the central node denotes the total number of items where $v > 0$ (i.e., $q > 1$, positively associated); 2) the length of the branch represents the value of v for that respective item; 3) the thickness of a branch represents the number of items with the same v at a given point; 4) the fluctuation of a branch conveys variation in v across different items in the same category corresponding to that branch.

The bacterium model helps users understand the visualized results by tapping in to prior knowledge about organic matter. For example, using the bacterium metaphor, the viewer compares multiple bacteria generated across different groups (e.g., younger and older females), as shown in Fig. 5. A bacterium that has longer branches would denote that the disease category placed in the central node has stronger effects on occurrences of the other diseases (e.g., an older female has higher chance to have some neoplasms after a mental disorder).

4.2 Generation Algorithm

In order to generate an organic appearance for each branch, our model employs L-systems [22, 30], which recursively create botanical shapes, lightnings, etc. Fig. 2 shows the first two steps of the recursion of our algorithm. Our algorithm starts from two points S and E . S is the point at $v = 0$ and E is the point where the maximum value of v for that branch is attained. The algorithm recursively decides the placement and values of the middle point M . Let p_i , v_i , w_i be a 2D coordinate, a value of v , and a width of a branch at point i (S , M , or E) respectively. p_M , v_M , w_M are calculated with

$$p_M = (p_S + p_E)/2 + \vec{u} \quad (2)$$

$$v_M = (v_S + v_E)/2 \quad (3)$$

$$w_M = \alpha N_{\leq v_M} \quad (4)$$

where $\vec{u} = c\beta\sigma\vec{n}$, and c is either 1 or -1 depending on whether the number of recursions at that point is an even or odd number. σ is the

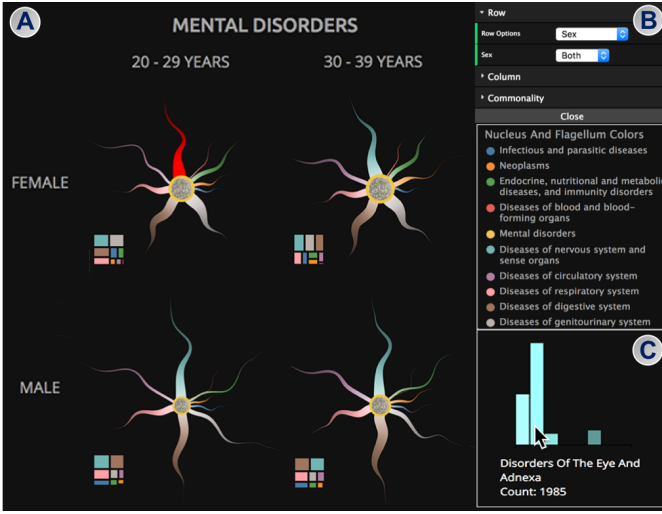


Fig. 5: The user interface of our application example.

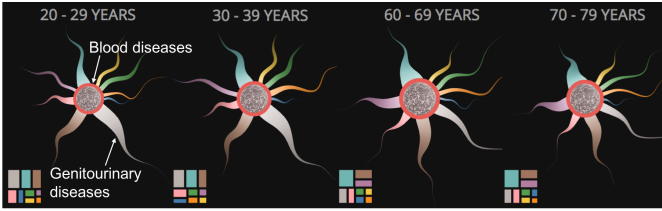


Fig. 6: The co-occurrences with the blood diseases in female groups.

are reported the most for females of these two age groups. We then explore how aging impacts the co-occurrences of diseases with blood diseases in females. We add the female age groups of 60–69 and 70–79, as shown in Fig. 6. We find that the genitourinary branches are much thinner and shorter when compared with the age groups of 20–29 and 30–39, indicating the co-occurrence of genitourinary diseases decreases in the older age groups. This finding is also supported by medical studies about this relationship between age and genitourinary diseases in females. At first, younger females are at a higher risk for anemia and autoimmune disorders [5, 31]. In addition, females in their 20–30’s have a higher risk to suffer from sexually transmitted infections, including UTI and STD [21, 23].

5.3 Case Study 2: Co-occurrences of Mental Disorders

In this second example, we explore the co-occurrence of mental disorders in males and females. We select mental disorders to be the central node of our bacteria. We then include males and females in the age ranges of 10–19 and 20–29. The visualized results are shown in Fig. 5. We can see that four bacterium metaphors have varying sizes and length in the central nodes and branches.

First, we examine a single bacterium for males of the age group 20–29. The flagellum for both nervous system (the branch colored with light blue) and digestive disorders (brown) stand out. Comparing the male 20–29 group to the female 20–29 group, we notice that the central node size for the female is larger. This indicates that this female group has more reported cases of other diseases co-occurring with mental disorders than with this male group. We also notice that unlike the male group, genitourinary diseases (gray) co-occurred more often as indicated with its wider width. Also, with all four bacteria, we can see that the central node size increases as age increases in both females and males.

As stated earlier, the length of a branch depends on the value of co-occurrence with the central node. Branch thickness is indicative of the number of diseases that have the same association. Therefore, in all four bacteria, we see that mental disorders (represented with the central nodes) and digestive diseases (brown branches) have a strong association. This finding is reasonable from our everyday experience

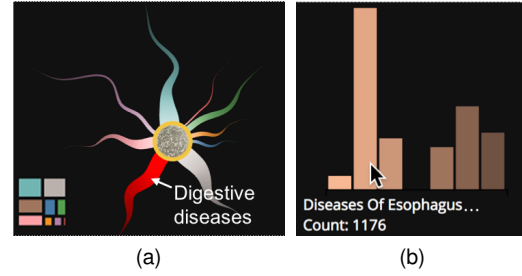


Fig. 7: Exploring the diseases' subcategories from the co-occurrence of the female 20–29 group visualized in Fig. 5. In (a), the branch for digestive diseases is selected. Its subcategories are visualized in (b).

of stomach pains caused by our anxiety.

We can also show more detailed information of the co-occurrence of diseases by selecting each branch. For example, in Fig. 5A, we select the branch for nervous system diseases for the female age group of 20–29, as highlighted with high-saturated red. Then, in Fig. 5C, the system visualize the number of diseases included in the subcategories of the nervous system diseases. As shown in Fig. 5C, when hovering the mouse over the longest bar, we can see that the dominant group is “disorders of the eye and adnexa”. Similarly, as shown in Fig. 7a, we select the branch for digestive diseases and find that the subcategory for “diseases of esophagus, stomach, and duodenum” in Fig. 7b is also highly associated with mental disorders. Pulling all these details together, we can ask why patients with mental disorders also report either diseases that fall under “diseases of esophagus, stomach, and duodenum”, or “disorders of the eye and adnexa”. Our application enables us to ask such questions, which could be the starting point of additional inquiry or research. Interestingly, some researchers have found relationships between irritable bowel syndrome and anxiety [12] as well as evidence showing those who go blind also tend to experience anxiety or forms of depression [26].

6 CONCLUSIONS AND FUTURE WORK

The use of a visual metaphor is effective when presenting complex information to the public. It provides a stepping stone for them to engage and learn more of an unfamiliar and complex topic. We have introduced an organic metaphor model to assist the public in interpreting conditional co-occurrences. Through a set of case studies, we demonstrated how we can uncover interesting findings with our application example, using our model.

The intent of our visual metaphor is to provide the general public a method on how view conditional co-occurrence data. We consider that traditional methods either focus on one disease and its co-occurrences or present all disease associations. As demonstrated in the case studies, our model provides an overview of these associations, and can be used to explore a variety of relationships when placed within a system. However, our model would not scale well if there are many item categories. Also, in both of case studies, we visualize only a limited number of groups as branches. We utilized the hierarchical categories of the dataset in order to summarize co-occurrence information into a small number of categories. Adding animation to our organic metaphor is one interesting future direction. In order to better attract the public, we could employ effective organic animation.

For future iterations, we plan to conduct user studies to evaluate our metaphor in more depth. We want to evaluate the impact of attractiveness on both user’s engagement for data exploration and abandonment rate for complex tasks. Also, we would like to verify this effect on different types of audience: domain experts, higher learners, and general public.

ACKNOWLEDGMENTS

The authors wish to thank both Dr. Shabbir Syed Abduland and Dr. Yu-Chuan Li at Taipei Medical University for the discussion on the dataset [27]. This research was supported in part by the U.S. National Science Foundations through grant IIS-1320229.

REFERENCES

- [1] N. Cawthon and A. V. Moere. The effect of aesthetic on the usability of data visualization. In *Information Visualization*, pp. 637–648. IEEE, 2007.
- [2] P. Chintagunta and S. Haldar. Investigating purchase timing behavior in two related product categories. *Journal of Marketing Research*, 35(1):43–53, 1998.
- [3] J. Chung and V. R. Rao. A general choice model for bundles with multiple-category products: Application to market segmentation and optimal pricing for bundles. *Journal of Marketing Research*, 40(2):115–130, 2003.
- [4] CMS and NCHS. The international classification of diseases: 9th revision, clinical modification: ICD-9-CM, 1991.
- [5] B. J. Cohen and Y. Gibor. Anemia and menstrual blood loss. *Obstetrical and Gynecological Survey*, 35(10):597–618, 1980.
- [6] D. Constanza, R. Carlos, G. Lara, M. Nieves, P. Gemma, V. Sergi, T. Rosa, R. J. A., and C. Miguel. Attention deficit hyperactivity disorder in cocaine-dependent adults: A psychiatric comorbidity analysis. *The American Journal on Addictions*, 22(5):466–473, 2013.
- [7] K. Dasu, S. Bae, T. Fujiwara, and K.-L. Ma. Learning about disease associations in Taiwan. IEEE PacificVis 2018 Visual Storytelling Contest, <https://k-dasu.github.io>. Accessed: 2018-6-11.
- [8] F. G. Davis, B. J. McCarthy, S. Freels, V. Kupelian, and M. L. Bondy. The conditional probability of survival of patients with primary malignant brain tumors. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 85(2):485–491, 1999.
- [9] M. Divo, C. Cote, J. P. de Torres, C. Casanova, J. M. Marin, V. Pinto-Plata, J. Zulueta, C. Cabrera, J. Zagaceta, G. Hunninghake, et al. Comorbidities and risk of mortality in patients with chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, 186(2):155–161, 2012.
- [10] N. K. Dubey, S. Syed-Abdul, P. A. Nguyen, R. Dubey, U. Iqbal, Y.-C. Li, W.-H. Chen, and W.-P. Deng. Association between anxiety state and mitral valve disorders: A taiwanese population-wide observational study. *Computer Methods and Programs in Biomedicine*, 132:57–61, 2016.
- [11] A. R. Feinstein. The pre-therapeutic classification of co-morbidity in chronic disease. *Journal of Chronic Diseases*, 23(7):455–468, 1970.
- [12] M. D. Fossey and R. B. Lydiard. Anxiety and the gastrointestinal system. *Psychiatric Medicine*, 8(3):175–186, 1990.
- [13] F. N. Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2):238–246, 1980.
- [14] T.-L. Fung, J.-K. Chou, and K.-L. Ma. A design study of personal bibliographic data visualization. In *Pacific Visualization Symposium (PacificVis)*, pp. 244–248. IEEE, 2016.
- [15] T. L. Fung and K.-L. Ma. Visual characterization of personal bibliographic data using a botanical tree design. In *IEEE VIS Workshop on Personal Visualization: Exploring Data in Everyday Life*, vol. 15, 2015.
- [16] D. Holten. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on visualization and computer graphics*, 12(5):741–748, 2006.
- [17] C.-W. Huang, S. Syed-Abdul, W.-S. Jian, U. Iqbal, P.-A. A. Nguyen, P. Lee, S.-H. Lin, W.-D. Hsu, M.-S. Wu, C.-F. Wang, K.-L. Ma, and Y.-C. J. Li. A novel tool for visualizing chronic kidney disease associated polymorbidity: a 13-year cohort study in taiwan. *Journal of the American Medical Informatics Association*, 22(2):290–298, 2015.
- [18] G. A. Kaplan, S. A. Everson, and J. W. Lynch. The contribution of social and behavioral research to an understanding of the distribution of disease: a multilevel approach. In *Smedley and Syme. Promoting Health: Intervention Strategies from Social and Behavioral Research*, pp. 31–55. National Academy Press, 2000.
- [19] E. Karni and D. Schmeidler. Self-preservation as a foundation of rational behavior under risk. *Journal of Economic Behavior and Organization*, 7(1):71–81, 1986.
- [20] T. M. Koppie, A. M. Serio, A. J. Vickers, K. Vora, G. Dalbagni, S. M. Donat, H. W. Herr, and B. H. Bochner. Age-adjusted charlson comorbidity score is associated with treatment decisions and clinical outcomes for patients undergoing radical cystectomy for bladder cancer. *Cancer*, 112(11):2384–2392, 2008.
- [21] C. M. Kunin. Urinary tract infections in females. *Clinical Infectious Diseases*, 18(1):1–10, 1994.
- [22] C. LaPointe and D. Stiert. Volume lightning rendering and generation using l-systems. *Advanced Computer Graphics*, 2009.
- [23] W. C. Miller, C. A. Ford, M. Morris, M. S. Handcock, J. L. Schmitz, M. M. Hobbs, M. S. Cohen, K. M. Harris, and J. R. Udry. Prevalence of chlamydial and gonococcal infections among young adults in the united states. *JAMA*, 291(18):2229–2236, 2004.
- [24] M. Moldovan, R. Enikeev, S. Syed-Abdul, P. A. Nguyen, Y.-C. Chang, and Y. C. Li. Disease universe: Visualisation of population-wide disease-wide associations. *Advances in Systems Science and Applications*, 14(2):144–158, 2014.
- [25] M. A. Moni, H. Xu, and P. Li. CytoCom: a Cytoscape app to visualize, query and analyse disease comorbidity networks. *Bioinformatics*, 31(6):969–971, 2015.
- [26] M. M. Moschos. Physiology and psychology of vision and its disorders: a review. *Medical Hypothesis, Discovery and Innovation in Ophthalmology*, 3(3):83, 2014.
- [27] National Health Research Institutes. National Health Insurance Research Database, Taiwan. <https://nhird.nhri.org.tw/en/>. Accessed: 2018-7-29.
- [28] H. C. Purchase. Metrics for graph drawing aesthetics. *Journal of Visual Languages & Computing*, 13(5):501–516, 2002.
- [29] J. S. Risch. On the role of metaphor in information visualization. *arXiv preprint arXiv:0809.0884*, 2008.
- [30] G. Rozenberg and A. Salomaa. *The mathematical theory of L systems*, vol. 90. Academic press, 1980.
- [31] K. Rubtsova, P. Marrack, and A. V. Rubtsov. Sexual dimorphism in autoimmunity. *The Journal of Clinical Investigation*, 125(6):2187–2193, 2015.
- [32] G. Russell and A. Petersen. Analysis of cross category dependence in market basket selection. *Journal of Retailing*, 76(3):367–392, 2000.
- [33] A. Sallaberry, Y.-c. Fu, H.-C. Ho, and K.-L. Ma. Contact trees: Network visualization beyond nodes and edges. *PLoS one*, 11(1):e0146368, 2016.
- [34] J. F. Tellez-Zenteno, S. B. Patten, N. Jetté, J. Williams, and S. Wiebe. Psychiatric comorbidity in epilepsy: A populationbased analysis. *Epilepsia*, 48(12):2336–2344, 2007.
- [35] S. Wang, Y. Tanahashi, N. Leaf, and K.-L. Ma. Design and effects of personal visualizations. *IEEE Computer Graphics and Applications*, 35(4):82–93, 2015.
- [36] R. J. Williams, A. Howe, and K. S. Hofmockel. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Frontiers in Microbiology*, 5:358, 2014.
- [37] H. Yang, L. Rothrock, S. Kumara, and R. Singareddy. VisOSA: Visualizing obstructive sleep apnea symptoms and comorbidities. In *Industrial and Systems Engineering Research Conference*, p. 1582. Institute of Industrial and Systems Engineers (IISE), 2013.
- [38] C. Zhou, Y.-L. Wu, G. Chen, J. Feng, X.-Q. Liu, C. Wang, S. Zhang, J. Wang, S. Zhou, S. Ren, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced egfr mutation-positive non-small-cell lung cancer (optimal, ctong-0802): a multicentre, open-label, randomised, phase 3 study. *The Lancet Oncology*, 12(8):735–742, 2011.
- [39] N. Zhou, J. Saltz, and K. Mueller. Maps of human disease: A web-based framework for the visualization of human disease comorbidity and clinical profile overlay. In F. Wang, G. Luo, C. Weng, A. Khan, P. Mitra, and C. Yu, eds., *Biomedical Data Management and Graph Online Querying*, pp. 47–60. Springer, 2016.