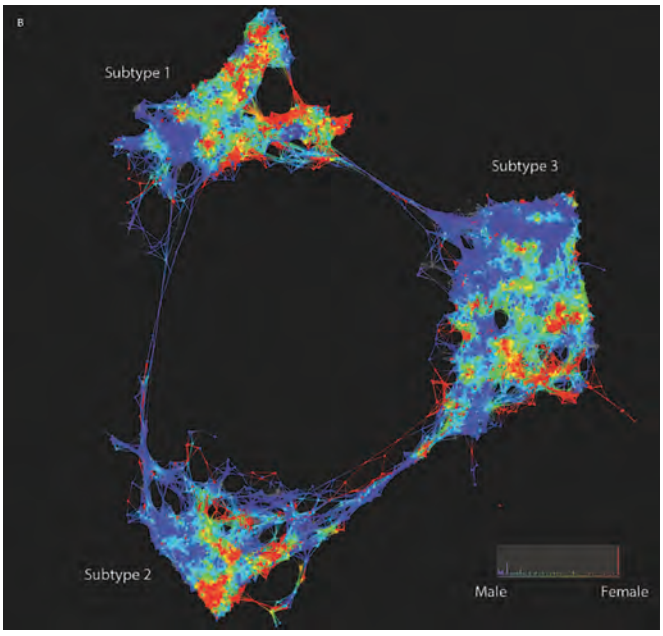

Figure 1. A network model for diabetics shows three distinct types, one heavily correlated with cancer.

Gunnar Carlsson

Topological Modeling of Complex Data

In the talk, we will be discussing the modeling of complex data sets by networks. Here are a couple of examples of applications of this idea. The first was a model for a data set of type 2 diabetics constructed by a group at the Mt. Sinai School of Medicine. The data set included genomic data as well as information from electronic medical records. Figure 1 shows a layout of the network that was obtained.

You can see that the network contains three large groups connected by "thin wires." The conclusion the Mt. Sinai researchers were able to draw from this is that type 2 diabetes, rather than being a single disease, is actually made up of three distinct types, and they found that one of these groups was heavily correlated with cancer. The finding that there are three distinct diseases will clearly have implications for treatment of the disease and constitute a contribution to "precision medicine."

A second example comes from the laboratory of David Schneider, a microbiologist at Stanford University. He studies the progression of infectious diseases. He has constructed a number of data sets using both physiological as well as genomic measurements on the subjects. Figure 2 shows an image of a collection of network models he constructed for mice and humans infected by flu and malaria.

You will notice that the models are both loops. This reflects the fact that they represent a phenomenon that begins at the healthy state, proceeds through gradual development of the disease until the immune response becomes strong, and then returns to the healthy state.

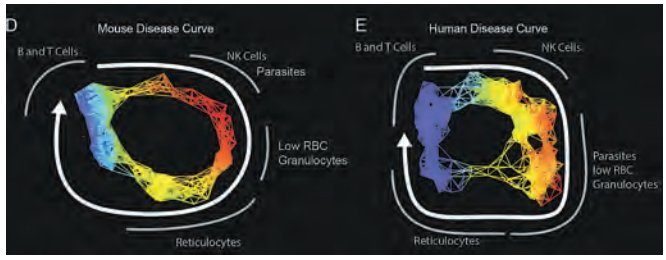

Figure 2. Mouse and human disease network models each form a loop through disease and recovery.

These models provide a time-independent model for the actual state of the subject. It is important to have such a model since (a) the progression of disease occurs at different rates for different subjects and (b) we don't generally have knowledge of the time the infection occurred. This should also be viewed as a contribution to precision medicine.

In the lecture we will discuss these and numerous other examples.

Image Credits

Figure 1 courtesy of "From identification of type 2 diabetes subgroups through topological analysis of patient similarity," L. Li, W-Y. Cheng , B. Glicksberg, O. Gottesman, R. Tamler, R. Chen, E. Bottinger, and J. Dudley *Science Translational Medicine* 28 Oct 2015: Vol. 7, Issue 311, pp. 311ra174. DOI:10.1126/scitranslmed.aaa9364.

Figure 2 courtesy of From Tracking Resilience to Infections by Mapping Disease Space, B.Y. Torres, J.H.M. Oliveira, Tate A. Thomas, P. Rath, K. Cumnock, and D.S. Schneider, *PLoS Biol.* **14**(4): e1002436. DOI:10.1371/journal.pbio.1002436.

Author photo courtesy of Gunnar Carlsson.

Gunnar Carlsson

ABOUT THE AUTHOR

Gunnar Carlsson spent most of his career working in the pure aspects of topology, but has turned to applications of the subject in the last fifteen years.

Gunnar Carlsson is professor of mathematics at Stanford University. His email address is carlsson@stanford.edu.

DOI: http://dx.doi.org/10.1090/noti1618