

New Computational Tools Help Solve Range of Population Biology Puzzles

Recent advances in sensing devices and other data-collection technologies are enabling scientists to gather huge amounts of information about animal populations and their interactions. Computer scientists at the University of Illinois at Chicago are developing new computational tools that could enable researchers to put that information to use in a wide variety of ways—from protecting endangered species to tracking terrorists.

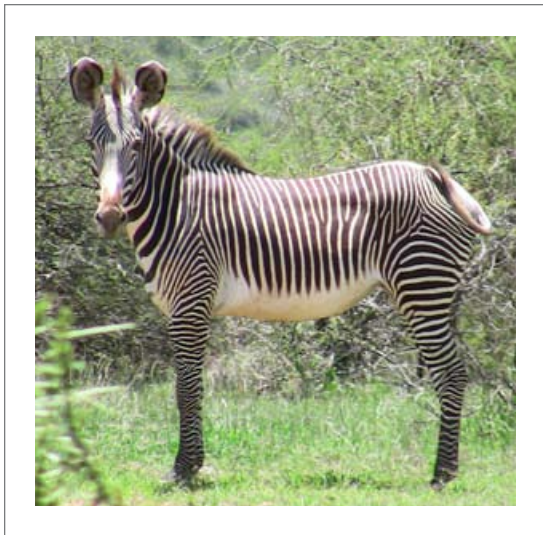
The Plains zebra and the Grévy's zebra both feed on the same grasses and occasional shrubs, wander much of the same semi-arid grasslands of western Africa and contend with the same predators. Yet for some reason, Plains zebra populations are strong while the Grévy's zebra is severely endangered.

Wildlife biologists believe that this has something to do with differences in how the animals within each species interact—especially in reacting to predators such as lions. Researchers who are using population biology methods to investigate this mystery are benefiting from recent breakthroughs in data-gathering technologies such as wireless GPS devices, which make it possible to continuously track the locations and interactions of individual animals.

“Unfortunately, biologists’ ability to analyze data seriously lags behind their ability to collect it,” says Tanya Berger-Wolf, a professor of computer science and director of the Laboratory for Computational Population Biology at the University of Illinois at Chicago.

Berger-Wolf is leading an effort to develop new computational methods and tools that will enable biologists to paint a more complete picture of the social networks of zebras. The software tools they are creating—which combine ideas from social network analysis, data mining and distributed computing—could help explain why the Grévy's zebra is struggling to survive.

But Berger-Wolf's algorithmic tools could have far broader uses in population biology, which is the study of populations of organisms. These could range from studying



A computer scientist at the University of Illinois at Chicago is leading an effort to develop new computational tools that enable researchers to better understand the social interactions among zebra populations in western Africa, which in turn could help explain why one species is thriving and another is endangered.

Fast Facts

Project Principals:

Tanya Berger-Wolf, assistant professor, Department of Computer Science, University of Illinois at Chicago

Daniel Rubenstein, professor, Department of Ecology and Evolutionary Biology, Princeton University

Jared Saia, assistant professor, Department of Computer Science, University of New Mexico

Siva Sundaresan, researcher, Princeton University and Denver Zoological Foundation

Ilya Fischhoff, researcher, Princeton University

Web Sites:

compbio.cs.uic.edu

compbio.cs.uic.edu/~tanya

Microsoft Technology:

SQL Server 2005

Microsoft External Research

The Microsoft External Research Division of Microsoft Research partners with academia, governments and industry to advance scientific and computer science research that addresses some of the world's most urgent and significant social and technological challenges. Along with providing research expertise, financial and technical support, and advanced technologies to enable groundbreaking projects worldwide, Microsoft External Research is committed to providing interoperable tools and services to support every stage of the research process. Efforts are focused in four research areas—including Health and Wellbeing, which is dedicated to developing technologies that advance healthcare and help people make better choices about their health.

Microsoft External Research
research.microsoft.com/collaboration

viruses and microbes to analyzing the behavior and interactions of other wild animals as well as humans. Besides helping in efforts to protect endangered species, for instance, the tools could significantly increase our understanding of everything from epidemics to terrorist networks.

The project was launched in 2006, with software, technical and financial support from Microsoft External Research. Berger-Wolf and a computer scientist from the University of New Mexico began working with a research team from Princeton University headed by Daniel Rubenstein, who has studied zebras for more than two decades and is doing extensive research into decision making by animals.



"The support I received from Microsoft got me started on a line of research that has taken me in exciting and completely unexpected directions."

—Tanya Berger-Wolf,
assistant professor of computer science,
University of Illinois at Chicago

In the initial phase of the project, the computer scientists developed tools to analyze existing data that Rubenstein's team had gathered mostly through visual observations of interactions among Plains and Grévy's zebra populations. Based on the results of these preliminary analyses, the team received funding from the National Science Foundation (NSF) to develop even more powerful computational tools.

The focus of the three-year NSF-funded project is to build data-analysis tools that will enable biologists to take advantage of rapid advances in sensor technologies such as GPS collars. The goal is to better understand the social networks of animals and to decipher patterns in how animal interactions affect their response to predators or other pressures, such as human land-use development or encroachments by domestic livestock.

Says Berger-Wolf: "Until very recently, we looked at these networks as sort of a static object—here are all of the individuals, and here are the lines connecting any two individuals that have been interacting at any point in the past."

Using data from visual observations of zebra interactions as well as data collected over a nine-month period from 10 GPS-collared zebras in Kenya and about 40 domestic and feral horses in the United States, Berger-Wolf's team developed tools that offer a far more dynamic picture of the timing and sequence of interactions. The tools, which are built partly on Microsoft® SQL Server® 2005, are enabling the researchers to identify key interaction patterns that Berger-Wolf says "no human eye could have seen in the data."

The zebra researchers have known that, for some reason, Grévy's zebras are more likely than Plains zebras to fall prey to lions. The new computational tools "are starting to clarify the

connections of how social dynamics affect susceptibility to lions and also to other disturbances," says Berger-Wolf.

Berger-Wolf hopes to use the tools to help investigate similar circumstances in which endangered species are struggling to coexist with human development or domestic livestock. "The long-term goal is to figure out how animal species, both wild and domestic, use the same habitat and how they're affecting each other," she says.

Berger-Wolf has tested the tools by using them to analyze e-mail exchanges at Enron Corporation, using records that were made public after the energy company was subpoenaed following an accounting fraud scandal. She also used images from the Internet Movie Database, an online database of movies, television shows and actors, to analyze interaction patterns among celebrities. And she analyzed histories of close-range connections of Bluetooth devices, using data from several studies conducted by researchers at the Massachusetts Institute of Technology and the University of Massachusetts.

Berger-Wolf says the tools will eventually help researchers across a wide range of disciplines figure out which social connections, patterns or sequences are most important. The tools will make it easier to predict future interactions and patterns within social networks and show how behavior spreads through a population.

For instance, just as the zebra researchers have used the tools to identify "influential individuals"—the zebras that others are most likely to follow when fleeing a lion—business marketers could use the tools to figure out what sort of person will be most influential in getting other people to buy certain products.

The tools have numerous other possible uses, such as tracking terrorist networks to anticipate their actions. Berger-Wolf was invited to demonstrate the tools at a workshop organized by the Institute for Defense Analyses, a federally funded organization that assists the U.S. government in addressing national security issues.

The tools, which are all open source, will eventually be combined with population genetics tools that Berger-Wolf and her colleagues are developing under a separate NSF-funded project. For her work on combining genetics and social interactions into a computational framework that will enable biologists to study populations in a completely new way, Berger-Wolf received the NSF CAREER Award in 2008.

"Computation has changed the way we study nature," Berger-Wolf says. "The support I received from Microsoft got me started on a line of research that has taken me in exciting and completely unexpected directions."

This document is for informational purposes only. MICROSOFT MAKES NO WARRANTIES, EXPRESS OR IMPLIED, IN THIS SUMMARY. © 2009 Microsoft Corporation. All rights reserved. Microsoft and SQL Server are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. The names of actual companies and products mentioned herein may be the trademarks of their respective owners.