Title: Scalable Parallel Computing of Large-Scale Graph Analytics for Big Data

Friday, 1/26/18 | 11am-12pm | West Hall, W105
Speaker: Dr. Maleq Khan

Abstract: Graph analytics plays a critical role in the analysis of data from diverse sources such as the Internet, social networks, computational biology, scientific simulations, and finance. Graph-based analysis of data offers valuable insights and can lead to the discoveries of hidden patterns in massive datasets. Traditional algorithms do not work well for big data. Moreover, such huge data may not fit in the memory of a single processing unit, and thus, require distributed systems where the data is distributed among multiple compute nodes. High performance computing platforms such as MPI, Spark and Hadoop, are now essential parts of big data analytics as they provide frameworks for parallel and distributed computing with thousands of processing cores. In this talk, I will present some of my recent work on studying dynamics on networks using graph analytics and developing scalable parallel algorithms for some problems in large-scale graph analytics. I will begin the talk with a study of using graph analytics in understanding disease dynamics on social contact networks and how some properties of a network affect the dynamics on the network. Then I will discuss some unique challenges in massive-scale graph analytics posed by big data and present some of my work on developing scalable distributed-memory parallel algorithms for generating random graphs, counting triangles, and subgraph analysis. These algorithms scale very well to a large number of cores (more than a thousand cores) and can work on graphs with billions of edges and vertices efficiently.

Bio: Maleq Khan is an Assistant Professor in the Department of Electrical Engineering and Computer Science at Texas A&M University–Kingsville. He received his Ph.D. in Computer Science from Purdue University in 2007 and then worked in the Biocomplexity Institute of Virginia Tech as a postdoc and as a research scientist before joining Texas A&M University–Kingsville in 2016. His research interests are big data analytics, high performance computing, parallel and distributed computing, data science and data mining. His research work received a best paper award (DISC 2006) and a best paper award nomination (SC 2016). Additional details about Dr. Khan and his work can be found at www.maleqkhan.net.