1.	Record Nr.	UNINA9910299892503321
	Autore	Wierzcho Slawomir
	Titolo	Modern Algorithms of Cluster Analysis / / by Slawomir Wierzcho, Mieczyslaw Kopotek
	Pubbl/distr/stampa	Cham : , : Springer International Publishing : , : Imprint : Springer, , 2018
	ISBN	3-319-69308-5
	Edizione	[1st ed. 2018.]
	Descrizione fisica	1 online resource (XX, 421 p. 51 illus.)
	Collana	Studies in Big Data, , 2197-6503 ; ; 34
	Disciplina	519.53
	Soggetti	Computational intelligence
		Big data
		Applied mathematics
		Engineering mathematics
		Computational Intelligence
		Big Data
		Applications of Mathematics
		Big Data/Analytics
	Lingua di pubblicazione	Inglese
	Formato	Materiale a stampa
	Livello bibliografico	Monografia
	Nota di bibliografia	Includes bibliographical references and index.
	Sommario/riassunto	This book provides the reader with a basic understanding of the formal concepts of the cluster, clustering, partition, cluster analysis etc. The book explains feature-based, graph-based and spectral clustering methods and discusses their formal similarities and differences. Understanding the related formal concepts is particularly vital in the epoch of Big Data; due to the volume and characteristics of the data, it is no longer feasible to predominantly rely on merely viewing the data when facing a clustering problem. Usually clustering involves choosing similar objects and grouping them together. To facilitate the choice of similarity measures for complex and big data, various measures of object similarity, based on quantitative (like numerical measurement results) and qualitative features (like text), as well as combinations of the two, are described, as well as graph-based similarity measures for (hyper) linked objects and measures for

multilayered graphs. Numerous variants demonstrating how such similarity measures can be exploited when defining clustering cost functions are also presented.   In addition, the book provides an overview of approaches to handling large collections of objects in a reasonable time. In particular, it addresses grid-based methods, sampling methods, parallelization via Map-Reduce, usage of tree-structures, random projections and various heuristic approaches, especially those used for community detection.