

Zusammenfassung

Die grundlegende Thematik der vorliegenden Dissertation ist die Identifizierung dichter Regionen eines ungerichteten Graphen $G = (V, E)$. Eine dichte Region ist dabei eine Teilmenge der Knoten $V' \subset V$ mit vielen Kanten, die zwischen Knoten aus V' verlaufen, aber vergleichsweise wenigen Kanten zu Knoten in $V \setminus V'$. Die Bestimmung dieser Gebiete ist bei der Lösung des Graphpartitionierungsproblems (GPP) sowie verwandten Aufgaben der Clusteranalyse hilfreich. Das GPP besteht in der Erstellung einer Partitionierung von V in k gleich große Teilmengen (Partitionen, Cluster) derart, dass die Zahl der Kanten, die zwischen verschiedenen Clustern verlaufen, minimiert wird. Es gibt zahlreiche Anwendungen, die die Lösung dieser oder ähnlicher Fragestellungen benutzen. Beispiele sind unter anderem parallele numerische Simulationen, Netzwerkanalyse, Schaltkreisentwurf sowie Genanalyse in der Bioinformatik.

GPP und alle relevanten Formulierungen verwandter Partitionierungsprobleme sind \mathcal{NP} -schwer, so dass keine Polynomialzeit-Algorithmen für ihre optimale Lösung bekannt sind. Partitionierungsbibliotheken, die dem aktuellen Stand der Technik entsprechen, benutzen lokale Knotenaustauschheuristiken innerhalb eines mehrstufigen Verbesserungsprozesses. Sie erzielen damit gute Lösungen in sehr kurzer Zeit. Jedoch entsprechen diese Lösungen nicht in jedem Fall den Anforderungen der Benutzer. Dies betrifft zum einen die Wahl der Zielfunktion im Optimierungsprozess, zum anderen die Form der Partitionen. Außerdem sind die am häufigsten eingesetzten Partitionierungsheuristiken schwierig zu parallelisieren, da sie inhärent sequentielle Teile enthalten. Eine solche Parallelisierung ist aber notwendig für den effizienten Einsatz als Lastbalancierer in parallelen Anwendungen. Zur Clusteranalyse von Graphen, bei der die Partitionen bzw. Cluster nicht gleich groß sein müssen, gibt es kein Verfahren, das sowohl sehr effizient arbeitet, Ergebnisse von sehr hoher Qualität in vielen verschiedenen Anwendungen liefert, als auch theoretisch wohlverstanden ist.

Um diese Nachteile bestehender Methoden zu beseitigen, entwerfen und untersuchen wir in dieser Arbeit den gestörten Diffusionsprozess FOS/C. Er kann dichte Gebiete eines Graphen von solchen mit wenigen Kanten unterscheiden, was wir mit seiner Beziehung zu Random Walks erklären. Durch die Kombination von FOS/C mit BUBBLE – einem generischen Verfahren ähnlich zu Lloyds k -means-Algorithmus – erhalten wir den iterativen und inhärent parallelen Algorithmus BUBBLE-FOS/C zur (Re)Partitionierung und Clusteranalyse von Graphen. In unseren theoretischen Untersuchungen zu FOS/C und BUBBLE-FOS/C beleuchten wir den Bezug zu Random Walks und zur Pseudoinversen der Laplacematrix des Eingabagraphen. Die dabei erzielten Ergebnisse führen unter an-

derem zu einem verbesserten Lösungsprozess von FOS/C und zu einem Beweis, dass BUBBLE-FOS/C gegen ein lokales Optimum konvergiert, welches durch eine Potentialfunktion charakterisiert werden kann.

Da BUBBLE-FOS/C die Lösung vieler linearer Gleichungssysteme erfordert, konstruieren wir einen effizienten Löser auf Basis des algebraischen Mehrgitterverfahrens (AMG). Die Graphhierarchie, die von diesem Löser erstellt wird, benutzen wir gleichzeitig für den mehrstufigen Partitionierungsprozess, der lokale Verbesserungen mit BUBBLE-FOS/C durchführt. Obwohl unser AMG-Löser eine deutliche Beschleunigung im Vergleich zu vorherigen Implementierungen hervorruft, bleibt die Laufzeit von BUBBLE-FOS/C weiterhin sehr hoch. Daher kann die gute Lösungsqualität des Algorithmus, die in Experimenten zur Graphpartitionierung beobachtet werden kann, kaum in der Praxis genutzt werden. Weitere Möglichkeiten zur Beschleunigung werden diskutiert, aber sie sind entweder nicht immer erfolgreich oder erfordern eine sehr aufwändige Implementierung.

Deshalb entwickeln wir in einem nächsten Schritt eine sehr viel schnellere und einfachere Methode zur Verbesserung von Partitionierungen. Diese Methode basiert auf einem anderen gestörten Diffusionsverfahren, das nur begrenzte Bereiche des Graphen betrachtet und auch einen hohen Grad an Parallelität aufweist. Das neue Verfahren kombinieren wir mit BUBBLE-FOS/C zu einem neuen mehrstufigen heuristischen Algorithmus namens DIBAP. Eine Besonderheit dieser Kombination ist, dass ihre Mehrstufen-Hierarchie durch zwei verschiedene Konstruktionsansätze erstellt wird. Verglichen mit BUBBLE-FOS/C, zeigt die neue Heuristik eine deutliche Beschleunigung und erhält gleichzeitig die positiven Eigenschaften des langsameren Algorithmus. Ausführliche Experimente zeigen ein extrem gutes Verhalten bei der Partitionierung von Graphen, die aus numerischen Simulationen stammen. DIBAP erzeugt durchgängig bessere Ergebnisse als die sehr häufig eingesetzten Bibliotheken METIS und JOSTLE. Weiterhin haben wir mit unserem neuen Algorithmus eine große Zahl der besten bekannten Partitionierungen von sechs weit verbreiteten Benchmark-Graphen verbessert. Auch in den verwandten Problemen der Lastbalancierung durch Repartitionierung und der Clusteranalyse verbessert DIBAP die Lösungsqualität des Stands der Technik in vielen Fällen.

Insofern besteht die vorliegende Arbeit aus praktischen und theoretischen Fortschritten für die (Re)Partitionierung und die Clusteranalyse von Graphen durch die Entwicklung neuer erfolgreicher heuristischer Algorithmen und die theoretische Analyse einiger wichtiger Eigenschaften dieser Algorithmen.