



دانشگاه صنعتی امیرکبیر
دانشکده مهندسی کامپیوتر

گزارش تکلیف اول درس الگوریتم‌های شبکه‌های پیچیده

دانشجویان:

سید احمد نقوی نوزاد

ش-د: ۹۴۱۳۱۰۶۰

مهدی مهرتاش

ش-د: ۹۴۱۳۱۰۴۹

استاد:

دکتر امیرحائری

جواب سوال ۱

توضیح روش پیشنهادی برای تشخیص انجمن‌های با هم‌پوشانی

در این جا برای تشخیص انجمن‌های با هم‌پوشانی از Clique Percolation Method (CPM) استفاده می‌نمائیم. در این روش برآنیم که زیرمجموعه‌هایی از گره‌ها را بیابیم که تعداد یال‌های بسیاری میان آن‌ها وجود دارد و برای این کار از ایده‌ی clique یا همان زیرگراف کامل استفاده می‌کنیم. بدین منظور برای شروع کار باید پارامتر ورودی k که همان اندازه‌ی clique و به عبارتی تعداد گره‌های آن می‌باشد را داشته باشیم که در این مسئله‌ی خاص ما آن را ۳ در نظر گرفته‌ایم. سپس تمامی clique‌های با اندازه‌ی ۳ را در گراف مربوطه می‌یابیم و به دنبال آن گراف جدیدی را با استفاده از clique‌های حاصله می‌سازیم که گره‌های آن گراف، همان clique‌های یافت‌شده می‌باشند و دو clique با هم مجاورند اگر ۲ گره‌ی مشترک داشته باشند. حال در گراف جدید که clique graph نامیده می‌شود، هر مؤلفه‌ی همبند آن حکم یک انجمن را دارد و به عبارت دیگر اعضای انجمن از اجتماع اعضای clique‌های موجود در آن مؤلفه‌ی همبند به دست می‌آید. لازم به ذکر است که ممکن است انجمن‌های به دست آمده با یکدیگر همپوشانی داشته باشند و نیز رأس‌هایی باقی بمانند که به هیچ انجمنی تعلق ندارند.

در نهایت پس از حصول انجمن‌های گراف از معیار Modularity برای ارزیابی کیفیت و یا قدرت تقسیم شبکه به انجمن‌ها استفاده می‌کنیم.

جواب سوال ۲

نمایش خوشه‌های حاصله با استفاده از هر کدام از ۴ روش مسئله

- روش اول (Spectral Analysis):

در این روش با در نظر گرفتن مقدار صفر برای threshold مربوطه تنها به کسب دو خوشه اکتفا نمودیم که به قرار زیر می‌باشند:

Graph Dolphins	Group 01	2, 4, 5, 6, 7, 10, 13, 14, 15, 17, 18, 20, 23, 26, 29, 33, 36, 38, 40, 41, 42, 49, 50, 51, 52, 54, 55, 57, 58, 59, 61, 62, 64, 65, 69, 74, 75, 76, 79, 83, 84, 85, 86, 87, 91, 93, 94, 96, 97, 98, 99, 100, 103, 104, 107, 110, 111, 112, 115, 116, 118, 119, 120
	Group 02	1, 3, 8, 9, 11, 12, 16, 19, 21, 22, 24, 25, 27, 28, 30, 31, 32, 34, 35, 37, 39, 43, 44, 45, 46, 47, 48, 53, 56, 60, 63, 66, 67, 68, 70, 71, 72, 73, 77, 78, 80, 81, 82, 88, 89, 90, 92, 95, 101, 102, 105, 106, 108, 109, 113, 114, 117, 121
Graph Karate	Group 01	1, 2, 4, 5, 6, 7, 8, 11, 12, 13, 14, 17, 18, 20, 22
	Group 02	3, 9, 10, 15, 16, 19, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34

- روش دوم (Latent Space Models):

در این روش نیز برای تعیین تعداد بزرگترین مقادیر ویژه‌ی بهینه از ماتریس \tilde{P} از رابطه‌ی معروف به انرژی به صورت زیر استفاده می‌کنیم (اگر مقادیر ویژه را به صورت نزولی مرتب کنیم):

$$\frac{\lambda_1 + \lambda_2 + \lambda_3 + \dots}{\sum \lambda_i} \geq 0.8$$

در نهایت پس از حصول ماتریس S برای تعیین تعداد انجمن‌های بهینه مورد استفاده در الگوریتم k -means آرایه‌ای درست می‌کنیم که در آن ازای هر تعداد انجمن دلخواه میزان $modularity$ حاصله را قرار می‌دهیم و در هر مرحله بررسی می‌کنیم که آیا مقدار $modularity$ بیشینه شده است یا نه؛ مقدار معادل با بیشینه‌ی $modularity$ همان تعداد انجمن‌های بهینه را به دست می‌دهد:

Graph Name	Best Top EigenValues	Best Community No. for K-Means	Community Members
Dolphins	7	3	1, 3, 4, 5, 9, 11, 12, 13, 15, 16, 17, 19, 21, 22, 24, 25, 29, 30, 31, 34, 35, 36, 38, 39, 43, 44, 45, 46, 47, 48, 50, 51, 52, 53, 54, 56, 59, 60, 62, 64, 78, 84, 89, 97, 100, 101, 104, 106, 121
			37, 41, 63, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 79, 80, 81, 82, 85, 86, 87, 88, 92, 93, 94, 95, 96, 98, 99, 102, 103, 105, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120
			2, 6, 7, 8, 10, 14, 18, 20, 23, 26, 27, 28, 32, 33, 40, 42, 49, 55, 57, 58, 61, 83, 90, 91
Karate	3	3	15, 16, 19, 21, 23, 24, 27, 30, 33, 34
			3, 9, 10, 14, 20, 25, 26, 28, 29, 31, 32
			1, 2, 4, 5, 6, 7, 8, 11, 12, 13, 17, 18, 22

- روش سوم ($Affiliation$ -Graph Models):

در این روش نیز برای تعیین تعداد انجمن‌های بهینه از همان آرایه‌ی بیان‌شده در روش قبل استفاده می‌کنیم:

Graph Name	Best Community No. for Community Membership Strength Matrix F	Community Members
Dolphins	5	1, 3, 11, 31, 43, 44, 47, 48, 54
		16, 19, 22, 25, 30, 46, 52
		9, 15, 17, 21, 34, 35, 38, 39, 41, 44, 51, 53
		6, 7, 8, 10, 14, 18, 20, 42, 55, 58
		2, 18, 24, 26, 27, 28, 37, 4, 5, 12, 13, 23, 29, 32, 33, 36, 40, 45, 49, 50, 56, 57, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121
Karate	3	1, 2, 3, 4, 8, 14
		1, 5, 6, 7, 11
		24, 32, 33, 34, 9, 10, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31

- روش چهارم (CPM):

انجمن‌های حاصله از این روش نیز به شرح ذیل می‌باشند:

Graph Name	Communities No.	Community Members
Dolphins	4	1, 4, 9, 15, 16, 17, 19, 21, 22, 24, 25, 30, 34, 35, 37, 38, 39, 41, 44, 45, 46, 51, 52, 53, 60
		2, 6, 7, 8, 10, 14, 18, 20, 31, 33, 42, 55, 58

		1, 3, 9, 11, 21, 29, 31, 43, 48
		2, 18, 26, 27, 28
Karate	3	1, 2, 3, 4, 8, 9, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 27, 28, 29, 30, 31, 32, 33, 34
		1, 5, 6, 7, 11, 17
		25, 26, 32

جواب سوال ۳

گزارش معیارهای Modularity و NMI برای روش‌های مربوطه

Method Name	Modularity		NMI	
	Dolphins	Karate	Dolphins	Karate
Spectral Analysis	.1187	.3600	-	-
Latent Space Models	.3521	.3350	-	-
Affiliation-Graph Models	.4759	.3715	.5010	.6460
Clique Percolation Method	.3218	.2202	-	-

همانطور که مشاهده می‌گردد مقدار *modularity* برای هر دو گراف حاصله از روش *Affiliation-Graph Model* از بقیه بالاتر می‌باشد و علت این امر نیز می‌تواند این باشد که چون ما برای گراف‌های مدنظر یک مدل اولیه فرض نمودیم و سپس با تشکیل آرایه‌ای که قید شد به دنبال یافتن مقدار بهینه برای تعداد انجمن‌ها بودیم، لذا موفق به یافتن مدلی برای گراف‌ها شدیم که بهترین حالت تقسیم‌بندی به انجمن‌ها را به دست می‌دهد. لازم به ذکر است که برای محاسبه‌ی *NMI* برای روش *Affiliation-Graph Models* خوشه‌بندی اولیه و به قولی صحیح را خوشه‌بندی حاصل از روش *Latent Space Model* در نظر گرفتیم.

در روش *AGM* در ابتدا با در نظر گرفتن یک مدل اولیه برای گراف یعنی تعداد انجمن‌های اولیه و نیز مقادیر اولیه برای *Community Membership Strength Matrix* و نیز استفاده از روش نزول گرادیان، به دنبال بیشینه کردن میزان درست‌نمایی احتمال رخداد گراف اصلی از روی گراف فرضی (اولیه) هستیم و این امر می‌تواند در عین پیچیدگی محاسباتی بالایی که دارد مانند این مسئله ما را به جواب بهینه و شاید بهترین جواب از بین کلیه جواب‌ها رهنمون شود.

اما در روش *CPM* همانطور که در پاسخ سوال اول قید شد به دنبال یافتن زیرمجموعه‌هایی از گره‌ها هستیم که تعداد اتصالات بسیاری میان آن‌ها برقرار است و بدین منظور از مفهوم *clique* که همان زیرگراف کامل می‌باشد استفاده می‌نمائیم. چرا که این مسئله می‌تواند سبب یافتن زیرمجموعه‌هایی از گراف شود که بیشتر به مفهوم انجمن نزدیک هستند در عین اینکه شاید این روش موفق به یافتن همه‌ی انجمن‌ها نگردد. لازم به ذکر است که بار محاسباتی این روش بنا به اندازه‌ی اولیه‌ی *clique* می‌تواند بسیار بالا بوده و به عبارتی مسئله از نوع مسائل *NP-hard (NonDeterministic Polynomial Time Hard)* به حساب آید.