1.网上找各种现实网络的数据库2.设计能给定网络，给定头信息占比变化函数（fhead）（或者不好算就直接固定或分段处理），算冗余的方法3.设计能给定网络，从网络中抽象出几个特征参数（类似PCA）的算法，参数有独立性（正交性），完备性，与可解释性。4.机器学习（或其他方法）学习随着网络特征参数的变化，最优fhead是什么。

5.通过简单图检验，去增加解释性，以及通过其他数据库，检验拟合效果。

6.同时设计对照算法（指实验失败品），证明最终结果的优越性。

关于最小化函数

如果目标点确定，希望最小化的value\_min，应当是 平均1接受者1bit占用的宽带数

Value\_try\_max(指单位路径占用的送达的有效用户数) = (infogood\*sendnum)/(roadnumall\*infoall)

Value\_min1= (roadnumall\*infoall)/ (infogood\*sendnum)

Infogood + infohead = infoall

单次send的要求为：发出一棵以A为根节点的树，除A外有infohead个节点用户收到信息，包含信息infogood。

又或者，最小化冗余信息数与路径损失信息数。这就是指，Value\_min2 = (roadnumloss\*infogood)+sum(roadnum\*infohead)

同时又有闲着无聊的恒等式sum roadnum = sum pointedsend = sendall + sum(passeduser\*passtime)

sum(passeduser\*passtime) = roadnumloss + sum(passeduser\*passtime)(using new road)

在没有单纯途径点的情况下，即sum(passeduser\*passtime)(using new road)=0时，roadnumloss + sendnum = sum roadnum。在这时，尝试证明在平均意义下，V1与V2具有相同评价效果。这是因为sendnum固定，vsimple= roadnumall/infogood。当所有infohead相同时，v2=(roadnumall-sendall)\*infogood + roadnumall \* infohead = roadnumall \* infoall – sendall \* infogood = infogood\*(vsimple -sendall)

重新定义V2为冗余的量纲，即V2/=infogood， 那么就有v2=vsimple-sendall。

凑数成功！

6 4 4 3 v1=(6+4+4+3)\*x/((x-6)\*6+(x-4)\*4+(x-4)\*4)+(x-3)\*3)=17x/(17x-77)=x/(x-4.53) , x为容量

V2=(36+16+16+9)=77

5 5 5 4 v1= (5+5+5+4)\*x/((x-5)\*5\*3+(x-4)\*4)=x/(x-4.79)

V2=(25+25+25+16)=91

5 5 4 4 v1= (5+5+4+4)\*x/((x-5)\*5\*2+(x-4)\*4 \* 2)=18x/(18x-82)=x/(x-4.56)

V2= 1\*(x-5)+(25+25+16+16)=77+x

无论是哪个函数，第一种方法6 4 4 3都胜利。当x=15时，平均每个有效用户的路径占用为1.43. x=20时为1.29.x越大，该方案越趋于1.

冗余路径为0.头信息损失77，平均信息损失4.53.

Task3

1. 图的结构：链式，星形，完全图，大图（幂率分布）
2. 估值函数：请使用v1
3. 目的：1.全场广播2.对目标点广播——同时需要抽取目标点的特征。1可看作为2的特殊（容易计算）的情形。
4. 给定fhead算v1：需要找到一个较优的路径选择方案。注意到在以上各情况下，infoall与sendnum都为定值，所以只要最小化vsimple = roadnumall/（infoall-infohead）就好。同时认为各目标点差别不大，所以各包的infohead相同（平均意义上）。
5. infoall与图的（以A为出发点的，不过以哪一点出发都一样）diameter的比值应该是个重要的值。Infohead与infoall的比值也是。因为，infohead至少需要为diameter的值才能传达。
6. 认为网络是由一个个集聚团分级链接的。要把信息传开来，A要把信息释放指与其链接的更高层的瓶颈，链接至问题范围要求的最高层瓶颈后再传下来发给低层网络。认为冗余是在两个地方产生。一个为A上传至中心节点。一个是中心节点下传至局域网。前者单路冗余极多，冗余影响道路少。后者单路冗余少，但冗余影响道路极多。在每一个局域网层，都存在这样两个现象。
7. 增加头信息可以消除节点下传至局域网的冗余（直接包含这整个局域网的信息，下传一次即可）问题是要增加到什么地步，即，增加到几级局域网的量。
8. 容易确定增加到的局域网的层级。因为局域网大小随其层级指数增长，到某个可观测的量级肯定就不能再提升层级了。这时headpercent达到了一个可观测的差不多的量级。下面所做的就是确定常数，即，传播多少个该层级局域网。
9. 不妨假设一次传播路径量为X，路径量等于头信息量X。/\*因为中心节点数目是较少的，主要用户都在局域网p内，所以认为用户接受者数也为X。\*/一个局域网到A的距离为D，高层局域网P有N个这样的局域网。任务是要传播完这个高级局域网P。设一次传播n个p。局域网内仍需传播距离为d。局域网内用户数为u开始算数。Vsimple= roadnumall/（infoall-infohead）=N/n\*(D+X\*d)/(infoall-X)，同时又n=X/u。vsimple=N\*u/X\*(D+X\*d)/(infoall-X).其中，D大概相比X\*d极小，infoall大概相比X极大。又D+d=Distance，u=N^d,Vsimple=(N^(d+1)\*Distance+(X-1)\*d)/(X\*(infoall-X))
10. 经计算知，一个包越能盛327个用户数。（ipv6 253，Ipv4 379 ipv6截止2020年5月普及率32%（据说是google数据） ）

Task3的模型大纲

对指定的所有目标点，for each 目标点，djistra（广搜）到A的路，进行最短路组抽取。

下面的讨论只针对被抽取过的所有线路构成的网络

局域网聚合：当意外地，高级局域网节点数目远小于低级局域网时，向上归并一层，是为高阶局域网聚合。这个归并的门槛大约为b=1？需要看效果和某种计算来考虑这个归并的条件，需要看。其二为，同阶局域网聚合。先发至一个局域网，再传向n个同阶局域网。优化的时候，对同距的两个选择，应当选择能更小度的路来自动增加同阶聚合性。而这样就好了。这是由于距离与网络尺度的同步性。

1. 通过广搜得出到A点的距离分布曲线。由此得出对各个节点有多少丢向下一层级。这是曲线分布中提出的参数。

提出b（i）=D(i+1)/D(i)的数列。

2.通过距离分布曲线选定一个由在该曲线基础下可变系数（如一个包包含的局域网数目和局域网层级）描绘的粗略跑路方案，得出一个由曲线唯一确定的vsimple

3.对vsimple取极小值，确定系数，细化跑路方案。最好得出可变系数针对曲线分布中提出某些参数的解析解。

4.确认优化效果：用这个模型跑task2，向评委说：“通过这个理论的确可以做出一个真正的不错的方案了（这个方案的优秀性在task2里已论证），产生了一个给图丢方案的流水线。”

5.进一步流程化该流水线。

6.在特殊图上进行检验（星形图，链式图，完全图，幂率分布复杂网络）